

121/487

ACCACATLACCACTTCTAGTACCACCGTTACCAATTGTATTTGAT Majority
7160 7170 7180 7190 7200
7151 ACCACATTACCACTTTTCTACTCTAGTACCACCGTTACCAATTGTATTTGAT coh1_a12.seq
7151 ACCACATTACCACTTTTCTACTCTAGTACCACCGTTACCAATTGTATTTGAT a909_a12.seq
TGAGGTATCTTCTAATTTGATATCTCCTACTGGAATAATGACAGGTTTTA Majority
7210 7220 7230 7240 7250
7201 TGAGGTATCTTCTAATTTGATATCTCCTACTGGAATAATGACAGGTTTTA coh1_a12.seq
7201 TGAGGTATCTTCTAATTTGATATCTCCTACTGGAATAATGACAGGTTTTA a909_a12.seq
TGGTGATATTTTTATTAGCATCTGCTAAATGGGCGTCAATATCAATGGAA Majority
7260 7270 7280 7290 7300
7251 TGGTGATATTTTTATTAGCATCTGCTAAATGGGCGTCAATATCAATGGAA coh1_a12.seq
7251 TGGTGATATTTTTATTAGCATCTGCTAAATGGGCGTCAATATCAATGGAA a909_a12.seq
TCATATGGGTTATAAATTTTACCATTGTACCACCAGCCACGGAAACGATA Majority
7310 7320 7330 7340 7350
7301 TCATATGGGTTATAAATTTTACCATTGTACCACCAGCCACGGAAACGATA coh1_a12.seq
7301 TCATATGGGTTATAAATTTTACCATTGTACCACCAGCCACGGAAACGATA a909_a12.seq
GCCATCTGGCATTGTCTGGACGTCTCAGTAAGGCTGAATGGGAGCCATCGT Majority
7360 7370 7380 7390 7400
7351 GCCATCTGGCATTGTCTGGACGTCTCAGTAAGGCTGAATGGGAGCCATCGT coh1_a12.seq
7351 GCCATCTGGCATTGTCTGGACGTCTCAGTAAGGCTGAATGGGAGCCATCGT a909_a12.seq
CATAGGAATCAGGGTCAGTAGGCTCATTACTTGTGTAATTGCTGACCA Majority
7410 7420 7430 7440 7450
7401 CATAGGAATCAGGGTCAGTAGGCTCATTACTTGTGTAATTGCTGACCA coh1_a12.seq
7401 CATAGGAATCAGGGTCAGTAGGCTCATTACTTGTGTAATTGCTGACCA a909_a12.seq
GAAGCATCCAATGCTGGCTTTCCATCTGTACCAACAGCATCATTGCTGTA Majority
7460 7470 7480 7490 7500
7451 GAAGCATCCAATGCTGGCTTTCCATCTGTACCAACAGCATCATTGCTGTA coh1_a12.seq
7451 GAAGCATCCAATGCTGGCTTTCCATCTGTACCAACAGCATCATTGCTGTA a909_a12.seq
TATAATATGATAAATCTCCAGCCCTTTCCGCCAAATAGCTCTTAAATTGATAT Majority
7510 7520 7530 7540 7550
7501 TATAATATGATAAATCTCCAGCCCTTTCCGCCAAATAGCTCTTAAATTGATAT coh1_a12.seq
7501 TATAATATGATAAATCTCCAGCCCTTTCCGCCAAATAGCTCTTAAATTGATAT a909_a12.seq
CTTGAGTTACAGCACCTGAAAAGTTATAAGGTCTAATACTACCATCTGGA Majority
7560 7570 7580 7590 7600
7551 CTTGAGTTACAGCACCTGAAAAGTTATAAGGTCTAATACTACCATCTGGA coh1_a12.seq
7551 CTTGAGTTACAGCACCTGAAAAGTTATAAGGTCTAATACTACCATCTGGA a909_a12.seq
TTAACATAAATACCAACCGACTAATTTGTAAAGCGTCTTTTACGTACTTGTA Majority
7610 7620 7630 7640 7650
7601 TTAACATAAATACCAACCGACTAATTTGTAAAGCGTCTTTTACGTACTTGTA coh1_a12.seq
7601 TTAACATAAATACCAACCGACTAATTTGTAAAGCGTCTTTTACGTACTTGTA a909_a12.seq
CTTAGTTGTTGTATCAACATTTGAGAGACTAGTATCTGTCGTATAAATAGG Majority
7660 7670 7680 7690 7700
7651 CTTAGTTGTTGTATCAACATTTGAGAGACTAGTATCTGTCGTATAAATAGG coh1_a12.seq
7651 CTTAGTTGTTGTATCAACATTTGAGAGACTAGTATCTGTCGTATAAATAGG a909_a12.seq
CATCTTTAGTTGAGTCTGGGATCTTTATCTCGTGAATCATACTTATAATAA Majority
7710 7720 7730 7740 7750
7701 CATCTTTAGTTGAGTCTGGGATCTTTATCTCGTGAATCATACTTATAATAA coh1_a12.seq
7701 CATCTTTAGTTGAGTCTGGGATCTTTATCTCGTGAATCATACTTATAATAA a909_a12.seq
TATGTACCTGAAGCATCTTGGATATAAATCCCTTGTAATATCTGTATAATC Majority
7760 7770 7780 7790 7800
7751 TATGTACCTGAAGCATCTTGGATATAAATCCCTTGTAATATCTGTATAATC coh1_a12.seq
7751 TATGTACCTGAAGCATCTTGGATATAAATCCCTTGTAATATCTGTATAATC a909_a12.seq

FIGURE 21K

122/487

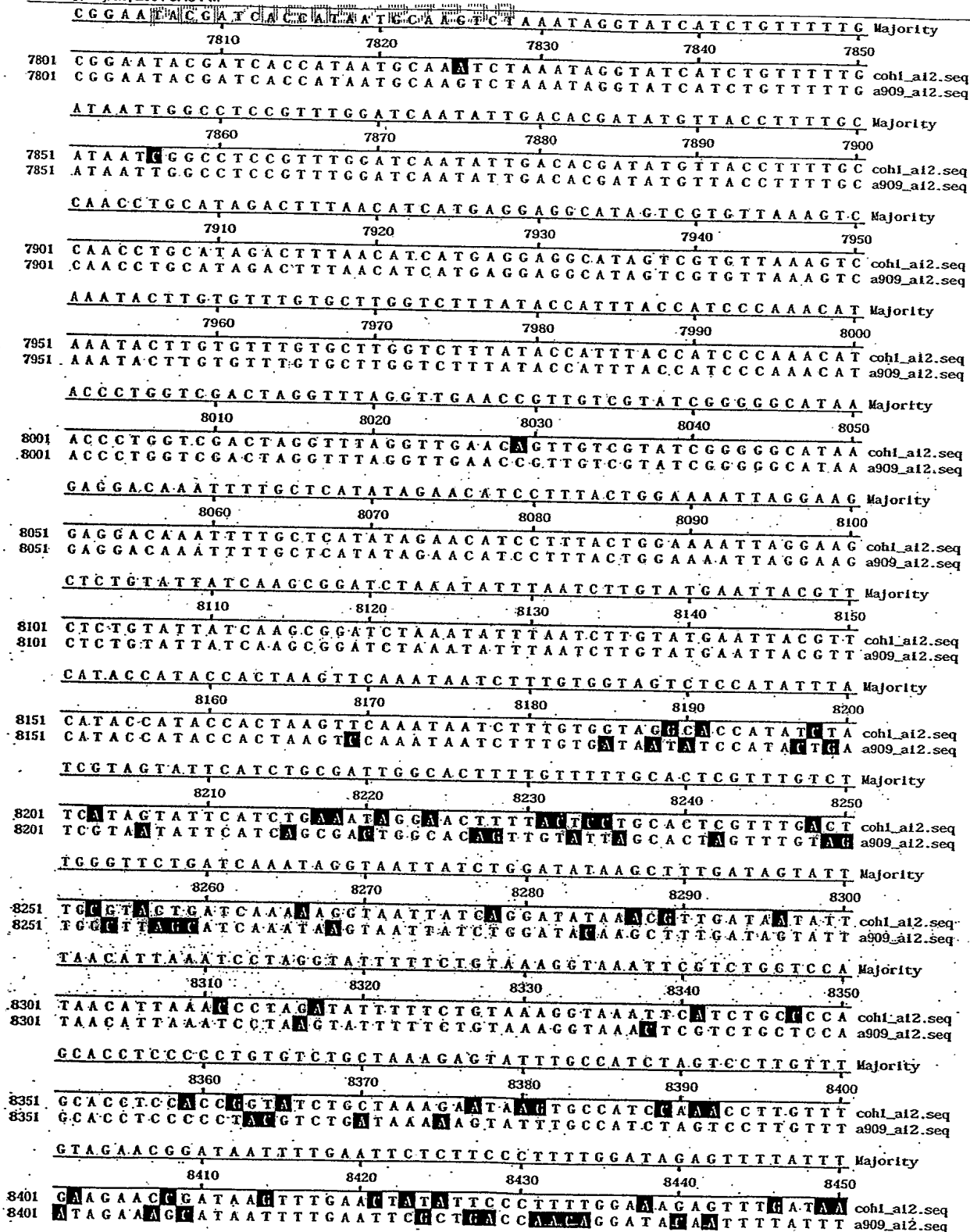


FIGURE 21L

123/487

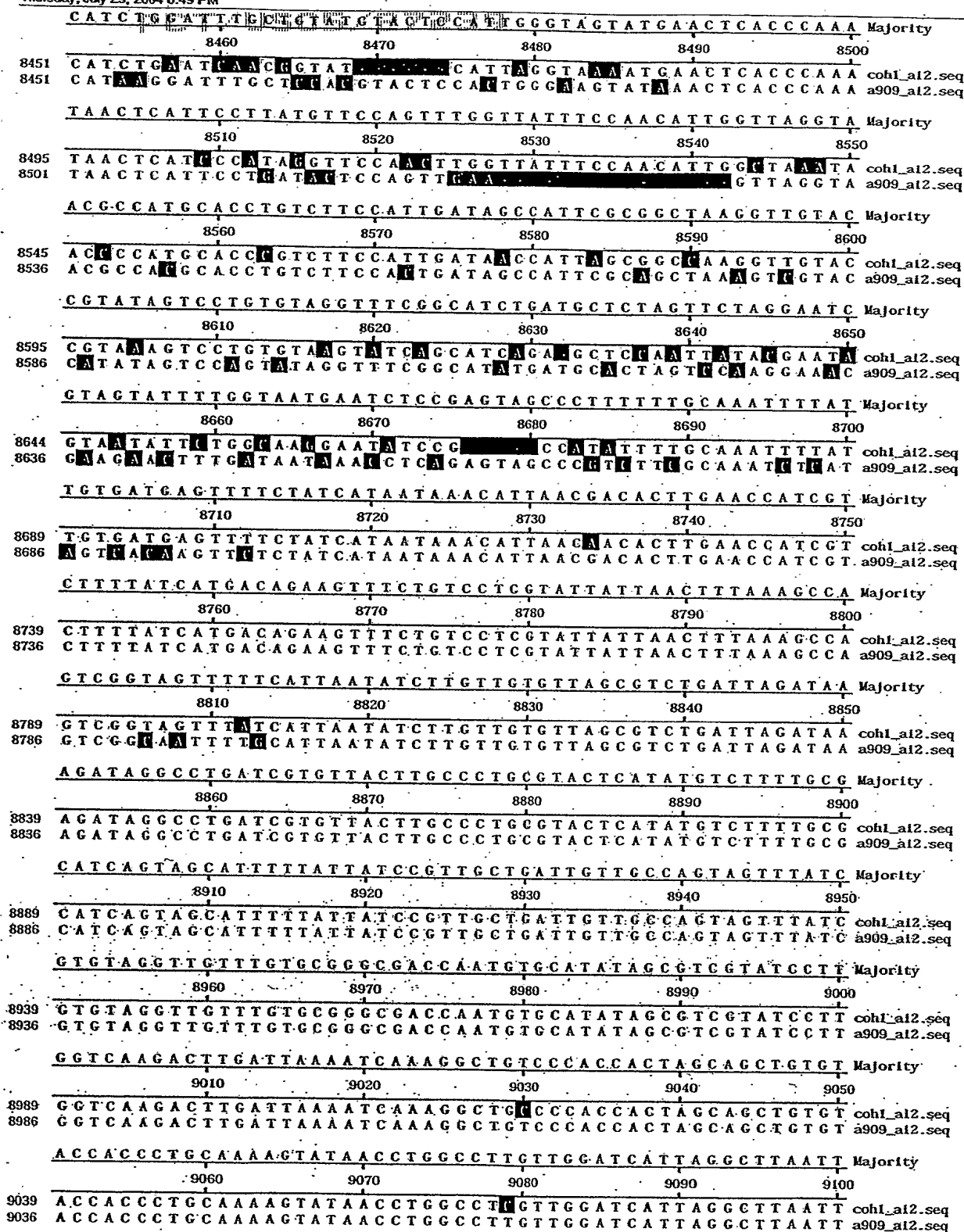


FIGURE 21M

124/487

	G T C C A A G C A G G T T G G T C T G T T A A C A C A C G A C G A G G T G C A A T A T A G G T A A C Majority									
	9110	9120	9130	9140	9150					
9089	G T C C A A G C A G G T T G G T C T G T T A A C A C A C G A C G A G G T G C A A T A T A G G T A A C coh1_al2.seq									
9086	G T C C A A G C A G G T T G G T C T G T T A A C A C A C G A C G A G G T G C A A T A T A G G T A A C a909_al2.seq									
	T C C T G T T G A T A A G T C G C C T G T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A Majority									
	9160	9170	9180	9190	9200					
9139	T C C T G T T G A T A A G T C G C C T G T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A coh1_al2.seq									
9136	T C C T G T T G A T A A G T C G C C T G T G T T G A A T T C A A C A C C G T A C G A T T C T T T A A a909_al2.seq									
	A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T Majority									
	9210	9220	9230	9240	9250					
9189	A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T coh1_al2.seq									
9186	A G A C A G G A T A A A G A C T T A T G T C T G A A G A T A C A G G T A A T G C T T G A A T T T C T a909_al2.seq									
	G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A Majority									
	9260	9270	9280	9290	9300					
9239	G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A coh1_al2.seq									
9236	G A T T C A G A A A G T G G A G C T C C A T T T T G A G T T T T A G A C C A A C C T A C A A A T A A a909_al2.seq									
	T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A G A T C C G C G G T T C Majority									
	9310	9320	9330	9340	9350					
9289	T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A A T C C G C G G T T C coh1_al2.seq									
9286	T A A T G T T G A A G T G G G T G G T G A A A C T T T A A A T G A A C T T A G A T C C G C G G T T C a909_al2.seq									
	C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A Majority									
	9360	9370	9380	9390	9400					
9339	C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A coh1_al2.seq									
9336	C G T T T C C T G A A C T A T C C A A T G G T A C T T G A T A A C G C T C C A A A A T A C T T T T A a909_al2.seq									
	C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G G Majority									
	9410	9420	9430	9440	9450					
9389	C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G G coh1_al2.seq									
9386	C C A G C A G A A T C A T C A T A T A A A G T T A C T G T C G C C A C T T T A C T C A T A A A T G G a909_al2.seq									
	A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T Majority									
	9460	9470	9480	9490	9500					
9439	A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T coh1_al2.seq									
9436	A C G A A C A T A A A T T T C T T T T G T C T C A G T T A C A G T T A T T G G C T C A C C A A A T T a909_al2.seq									
	T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A Majority									
	9510	9520	9530	9540	9550					
9489	T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A coh1_al2.seq									
9486	T A A C A G G G T C A C C A T A C T T T C C A G T A G T A G G A T C A T A G G T A T A C C A A C C A a909_al2.seq									
	T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C C A A C T T C T C C T A G A G A Majority									
	9560	9570	9580	9590	9600					
9539	T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C C A A C T T C T C C T A G A G A coh1_al2.seq									
9536	T T A A A A T G C T C T C C T G C T T T A A T C G T C G G A A T C C C A A C T T C T C C T A G A G A a909_al2.seq									
	T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A Majority									
	9610	9620	9630	9640	9650					
9589	T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A coh1_al2.seq									
9586	T T C T C C A T C T T T T A T A A T T T G A T G A T G A A C T T G C A T A C C T G A A G C T G T C A a909_al2.seq									
	G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A Majority									
	9660	9670	9680	9690	9700					
9639	G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A coh1_al2.seq									
9636	G G A A A T T A T A A T C A G T T C C G T C A T T A T T T T G A A A A T G G T A A G T T A A C C T A a909_al2.seq									
	G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C Majority									
	9710	9720	9730	9740	9750					
9689	G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C coh1_al2.seq									
9686	G G A A C T T C T G T A T T A T C C T C T T G A A C A A T T G C A T A A A T G G A G A A T G A A T C a909_al2.seq									

FIGURE 21N

125/487

	T	G	T	T	T	T	A	A	A	G	C	A	A	C	A	A	L	C	A	C	T	G	C	T	A	C	T	G	T	T	C	T	T	A	G	T	T	T	C	T	G	C	A	G	T	A	T	C	T	T		Majority																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
								9760																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	

FIGURE 210

FIGURE 21P

127/487

CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA Majority
11060 11070 11080 11090 11100
11039 CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA coh1_a12.seq
11035 CGCCTCTCTTATCTACTAAATTTCTTTACGTTTTCTAAATAGGTATCTGCA a909_a12.seq

GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC Majority
11110 11120 11130 11140 11150
11089 GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC coh1_a12.seq
11085 GCTAACTGAGCAGTCTCAAAAAATACCACTAAGCATTAAAGGAAGCTTCGGC a909_a12.seq

AATAGAACCAGCTTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT Majority
11160 11170 11180 11190 11200
11139 AATAGAACCAGCTTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT coh1_a12.seq
11135 AATAGAACCAGCTTTTGCTAATTTTGAATTTTTTATCATCTAGCGCTTCTT a909_a12.seq

TAAGTTGCTGAATCTCTTTCTCTTGTCTTTTCAATAAGAAGTTGCTGTTCT Majority
11210 11220 11230 11240 11250
11189 TAAGTTGCTGAATCTCTTTCTCTTGTCTTTTCAATAAGAAGTTGCTGTTCT coh1_a12.seq
11185 TAAGTTGCTGAATCTCTTTCTCTTGTCTTTTCAATAAGAAGTTGCTGTTCT a909_a12.seq

AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT Majority
11260 11270 11280 11290 11300
11239 AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT coh1_a12.seq
11235 AACATAAATTTCTAGCAAGTCTTTTCTTTTAAATTTTTTAAATCTTCCAT a909_a12.seq

CGCGATTACTTCTTAACTGAACCTTAAATTATCGTTTAGATATTATATC Majority
11310 11320 11330 11340 11350
11289 CGCGATTACTTCTTAACTGAACCTTAAATTATCGTTTAGATATTATATC coh1_a12.seq
11285 CGCGATTACTTCTTAACTGAACCTTAAATTATCGTTTAGATATTATATC a909_a12.seq

AAAGTTCTAACCCTTTAAACTCATTTTTTGTCTGTCTTTTTTCTCAAAAAA Majority
11360 11370 11380 11390 11400
11339 AAAGTTCTAACCCTTTAAACTCATTTTTTGTCTGTCTTTTTTCTCAAAAAA coh1_a12.seq
11335 AAAGTTCTAACCCTTTAAACTCATTTTTTGTCTGTCTTTTTTCTCAAAAAA a909_a12.seq

AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGA AAAATCTCATCGA Majority
11410 11420 11430 11440 11450
11389 AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGA AAAATCTCATCGA coh1_a12.seq
11385 AGTCTATGCTAAATTAACATTTTTGATAATTTTTTGA AAAATCTCATCGA a909_a12.seq

AGTCATTTTTCTTTTTGAAAGCTCGAATTCTAGGCATTAAAAAGCCATATA Majority
11460 11470 11480 11490 11500
11439 AGTCATTTTTCTTTTTGAAAGCTCGAATTCTAGGCATTAAAAAGCCATATA coh1_a12.seq
11435 AGTCATTTTTCTTTTTGAAAGCTCGAATTCTAGGCATTAAAAAGCCATATA a909_a12.seq

TC AAAATTGATATATGGCTTTTTTTATTATTTAAAAACAAAAGCAATCAATAG Majority
11510 11520 11530 11540 11550
11489 TC AAAATTGATATATGGCTTTTTTTATTATTTAAAAACAAAAGCAATCAATAG coh1_a12.seq
11476 TC AAAATTGATATATGGCTTTTTTTATTATTTAAAAACAAAAGCAATCAATAG a909_a12.seq

GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT Majority
11560 11570 11580 11590 11600
11539 GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT coh1_a12.seq
11492 GACAATAGCGTCAATTTAGTGACATAATCTATTACAGATTAAAGTTCTTTT a909_a12.seq

TGAATAATATAATCCAACTTTTCAACTGTTTTTTCCCATGTGAAATGTTT Majority
11610 11620 11630 11640 11650
11589 TGAATAATATAATCCAACTTTTCAACTGTTTTTTCCCATGTGAAATGTTT coh1_a12.seq
11542 TGAATAATATAATCCAACTTTTCAACTGTTTTTTCCCATGTGAAATGTTT a909_a12.seq

TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTTAATGCCTTAT Majority
11660 11670 11680 11690 11700
11639 TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTTAATGCCTTAT coh1_a12.seq
11592 TTTAATTCTTTTAGCAATATTCTGTTGTAGTTTCTCTCTTAATGCCTTAT a909_a12.seq

FIGURE 21Q

128/487

```

CTTTTAACTTAAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC Majority
11710 11720 11730 11740 11750
11689 CTTTACTAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC coh1_a12.seq
11642 CTTTACTAATAAATCAAGAGATTTCATGAGTGAAGTATTTTCTTCC a909_a12.seq

ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC Majority
11760 11770 11780 11790 11800
11739 ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC coh1_a12.seq
11692 ATGATGATTCTTAACCTCAGGGCTATCAATAAATTCAACTGTTCCACCGCC a909_a12.seq

ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG Majority
11810 11820 11830 11840 11850
11789 ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG coh1_a12.seq
11742 ATCTGTTGCAATAATAGCACTCGAAAGTAGACCAGCTTCTAAAAATAGAGG a909_a12.seq

TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT Majority
11860 11870 11880 11890 11900
11839 TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT coh1_a12.seq
11792 TTGGTAATCCCTCTCGGATACATTGATGGGTAAACAAAGATATCTGTCTGT a909_a12.seq

GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT Majority
11910 11920 11930 11940 11950
11889 GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT coh1_a12.seq
11842 GCCATTAAAGACATAGTCTGTTCAAAGTTTAAATTTCCCAAAAAAGTTAAT a909_a12.seq

CTGTTTGGAAGTATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC Majority
11960 11970 11980 11990 12000
11939 CTGTTTGGAAGTATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC coh1_a12.seq
11892 CTGTTTGGAAGTATATTTCTCTTTCAAATGTGCTAATTCAGGTCCGTCTC a909_a12.seq

CTGCAATCTGTAAATAAACAATTTTCAGAGTACTGTGACATCGAAAAATGCT Majority
12010 12020 12030 12040 12050
11989 CTGCAATCTGTAAATAAACAATTTTCAGAGTACTGTGACATCGAAAAATGCT coh1_a12.seq
11942 CTGCAATCTGTAAATAAACAATTTTCAGAGTACTGTGACATCGAAAAATGCT a909_a12.seq

TCTAAGAGCAATTCAATGCCCTTTTCTTTTAAATAATTCTACCAGCATAAGT Majority
12060 12070 12080 12090 12100
12039 TCTAAGAGCAATTCAATGCCCTTTTCTTTTAAATAATTCTACCAGCATAAGT coh1_a12.seq
11992 TCTAAGAGCAATTCAATGCCCTTTTCTTTTAAATAATTCTACCAGCATAAGT a909_a12.seq

GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTACGCAAAAAT Majority
12110 12120 12130 12140 12150
12089 GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTACGCAAAAAT coh1_a12.seq
12042 GATGAAAAATATCATCAGCAGATTTTTCAGGTAAGCCGTGTACGCAAAAAT a909_a12.seq

CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT Majority
12160 12170 12180 12190 12200
12139 CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT coh1_a12.seq
12092 CAGAGCCTAGACTTTTCAGATACCGAATTATAAAATAACTCTTTAGCTTCT a909_a12.seq

ATATTTAAATGTTTTAAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA Majority
12210 12220 12230 12240 12250
12189 ATATTTAAATGTTTTAAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA coh1_a12.seq
12142 ATATTTAAATGTTTTAAACCATTCAACGGCTTCTCTTGGATACCGCATAAAA a909_a12.seq

ATCTGGACGATAGTGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG Majority
12260 12270 12280 12290 12300
12239 ATCTGGACGATAGTGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG coh1_a12.seq
12192 ATCTGGACGATAGTGTCTTAACACGGCGCTGTGAGAGATGTTTCATAGATAG a909_a12.seq

CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA Majority
12310 12320 12330 12340 12350
12289 CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA coh1_a12.seq
12242 CTCCAAAGAAATCTAAAAAACGATTATTGACAGAAAAATGACTTGACCCA a909_a12.seq

```

FIGURE 21R

129/487

		12360	12370	12380	12390	12400	
		TGGTCTAAAAACAATACTAGGTAACGGTGGCTTCTTTGCAAAAGATAGCCC Majority					
12339		TGGTCTAAAAACAATACTAGGTAACGGTGGCTTCTTTGCAAAAGATAGCCC coh1_a12.seq					
12292		TGGTCTAAAAACAATACTAGGTAACGGTGGCTTCTTTGCAAAAGATAGCCC a909_a12.seq					
		12410	12420	12430	12440	12450	
		TTCTAGCGTTGTTAACTGAAAACGCTGATTACAAATCACAAAATCAATAT Majority					
12389		TTCTAGCGTTGTTAACTGAAAACGCTGATTACAAATCACAAAATCAATAT coh1_a12.seq					
12342		TTCTAGCGTTGTTAACTGAAAACGCTGATTACAAATCACAAAATCAATAT a909_a12.seq					
		12460	12470	12480	12490	12500	
		TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA Majority					
12439		TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA coh1_a12.seq					
12392		TTTCATCTGAAAACATGTTTCATCAGCGTGTGTTGTTCTCGATTTTTTGTTA a909_a12.seq					
		12510	12520	12530	12540	12550	
		ATAATAGGATAGCGCTGCTTGACAAAGTTTTTGGTCCGTAACCGGTAAAT Majority					
12489		ATAATAGGATAGCGCTGCTTGACAAAGTTTTTGGTCCGTAACCGGTAAAT coh1_a12.seq					
12442		ATAATAGGATAGCGCTGCTTGACAAAGTTTTTGGTCCGTAACCGGTAAAT a909_a12.seq					
		12560	12570	12580	12590	12600	
		TTTTCTACCCTTGCTTTTCATCTATAATCGGTAAATCATCATGATTAGTTG Majority					
12539		TTTTCTACCCTTGCTTTTCATCTATAATCGGTAAATCATCATGATTAGTTG coh1_a12.seq					
12492		TTTTCTACCCTTGCTTTTCATCTATAATCGGTAAATCATCATGATTAGTTG a909_a12.seq					
		12610	12620	12630	12640	12650	
		TTACAATAACAACACGGTAGCCACGCTTAACCAATCTGCTGTCATTTTA Majority					
12589		TTACAATAACAACACGGTAGCCACGCTTAACCAATCTGCTGTCATTTTA coh1_a12.seq					
12542		TTACAATAACAACACGGTAGCCACGCTTAACCAATCTGCTGTCATTTTA a909_a12.seq					
		12660	12670	12680	12690	12700	
		TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA Majority					
12639		TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA coh1_a12.seq					
12592		TCTGTATAACGTTCAATACCTCCGAGGAAGGGTAGATAATATCCTGAGAA a909_a12.seq					
		12710	12720	12730	12740	12750	
		AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT Majority					
12689		AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT coh1_a12.seq					
12642		AACAGCAACTGTTTTTACCTTATTTTCCATATTTTATCCACTTTTCATCAAT a909_a12.seq					
		12760	12770	12780	12790	12800	
		AAGCCATCTTTTAAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG Majority					
12739		AAGCCATCTTTTAAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG coh1_a12.seq					
12692		AAGCCATCTTTTAAAGCCTTTAATCATAGCAACTAATTTTTTGGCTCTTTTG a909_a12.seq					
		12810	12820	12830	12840	12850	
		CTCTTCTGCTACCAACACTCGAACAATTCATTTEGCGATAAAATACTAAAT Majority					
12789		CTCTTCTGCTACCAACACTCGAACAATTCATTTEGCGATAAAATACTAAAT coh1_a12.seq					
12742		CTCTTCTGCTACCAACACTCGAACAATTCATTTEGCGATAAAATACTAAAT a909_a12.seq					
		12860	12870	12880	12890	12900	
		ATTTGTGCGCGTTCTTTTACCATATTTTATTAATATAAATCGCATTC Majority					
12839		ATTTGTGCGCGTTCTTTTACCATATTTTATTAATATAAATCGCATTC coh1_a12.seq					
12792		ATTTGTGCGCGTTCTTTTACCATATTTTATTAATATAAATCGCATTC a909_a12.seq					
		12910	12920	12930	12940	12950	
		CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC Majority					
12889		CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC coh1_a12.seq					
12842		CGTATCATGTAATATTTTTCGAAATGGTGAATGATTCAATACATGAAAAAC a909_a12.seq					
		12960	12970	12980	12990	13000	
		ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC Majority					
12939		ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC coh1_a12.seq					
12892		ATGCCCAAATTTTTTAACTCGTGAAGAGTGTCCAATTTTCGTGTAAAAGAC a909_a12.seq					

FIGURE 21S

130/487

	CAATAAATAAATTAACCTGATAAGTCTTATATATCCCATCTCTGACAGACGATAA	Majority
	13010 13020 13030 13040 13050	
12989	CAATAAATAAATTAACCTGATAAGTCTTATATATCCCATCTCTGACAGACGATAA	cohl_al2.seq
12942	CAATAAATAAATTAACCTGATAAGTCTTATATATCCCATCTCTGACAGACGATAA	a909_al2.seq
	TTCATTTTCAGAGTCAACAAATCAATAAACATCTCTTCTGCAAAGCCAGA	Majority
	13060 13070 13080 13090 13100	
13039	TTCATTTTCAGAGTCAACAAATCAATAAACATCTCTTCTGCAAAGCCAGA	cohl_al2.seq
12992	TTCATTTTCAGAGTCAACAAATCAATAAACATCTCTTCTGCAAAGCCAGA	a909_al2.seq
	TGTTTCTTCAAAAACGCTCGTTTTTCATTAAAGCAGCCGAAGTAATACACT	Majority
	13110 13120 13130 13140 13150	
13089	TGTTTCTTCAAAAACGCTCGTTTTTCATTAAAGCAGCCGAAGTAATACACT	cohl_al2.seq
13042	TGTTTCTTCAAAAACGCTCGTTTTTCATTAAAGCAGCCGAAGTAATACACT	a909_al2.seq
	CTTCAATTTCTTTATAGTCAAAATTTCTTGCATCACTAAATCTTCACGGTTC	Majority
	13160 13170 13180 13190 13200	
13139	CTTCAATTTCTTTATAGTCAAAATTTCTTGCATCACTAAATCTTCACGGTTC	cohl_al2.seq
13092	CTTCAATTTCTTTATAGTCAAAATTTCTTGCATCACTAAATCTTCACGGTTC	a909_al2.seq
	ATATCTTGATACAAACAAGATAACATACCGACCTTAGGTAATGAAGGTA	Majority
	13210 13220 13230 13240 13250	
13189	ATATCTTGATACAAACAAGATAACATACCGACCTTAGGTAATGAAGGTA	cohl_al2.seq
13142	ATATCTTGATACAAACAAGATAACATACCGACCTTAGGTAATGAAGGTA	a909_al2.seq
	ATTTTCATAATTATCTATCAAAATCACCTAGGACAACCGAATCTTGATCTA	Majority
	13260 13270 13280 13290 13300	
13239	ATTTTCATAATTATCTATCAAAATCACCTAGGACAACCGAATCTTGATCTA	cohl_al2.seq
13192	ATTTTCATAATTATCTATCAAAATCACCTAGGACAACCGAATCTTGATCTA	a909_al2.seq
	AAGTCAAGAACCAATCAAAATTTCTTGCTACTGCAAATTGACCGATACAG	Majority
	13310 13320 13330 13340 13350	
13289	AAGTCAAGAACCAATCAAAATTTCTTGCTACTGCAAATTGACCGATACAG	cohl_al2.seq
13242	AAGTCAAGAACCAATCAAAATTTCTTGCTACTGCAAATTGACCGATACAG	a909_al2.seq
	TTCAAAGCATATGCAATTTCTTTATTTTCTGTTAAATAATCAACAGTTAG	Majority
	13360 13370 13380 13390 13400	
13339	TTCAAAGCATATGCAATTTCTTTATTTTCTGTTAAATAATCAACAGTTAG	cohl_al2.seq
13292	TTCAAAGCATATGCAATTTCTTTATTTTCTGTTAAATAATCAACAGTTAG	a909_al2.seq
	GTGCCCCCTCTTCATTATAATCGGCCACTAATTGAGAAATTTCTTCCTTAT	Majority
	13410 13420 13430 13440 13450	
13389	GTGCCCCCTCTTCATTATAATCGGCCACTAATTGAGAAATTTCTTCCTTAT	cohl_al2.seq
13342	GTGCCCCCTCTTCATTATAATCGGCCACTAATTGAGAAATTTCTTCCTTAT	a909_al2.seq
	TTTTCGAGCCATTATCTACCATGTAGATATGGCTTACTTGAGGATAAATT	Majority
	13460 13470 13480 13490 13500	
13439	TTTTCGAGCCATTATCTACCATGTAGATATGGCTTACTTGAGGATAAATT	cohl_al2.seq
13392	TTTTCGAGCCATTATCTACCATGTAGATATGGCTTACTTGAGGATAAATT	a909_al2.seq
	GCTCGAATGTTCTGATCTAAGCGTTCAATATTGCGGTTAAAGGTGACAAAT	Majority
	13510 13520 13530 13540 13550	
13489	GCTCGAATGTTCTGATCTAAGCGTTCAATATTGCGGTTAAAGGTGACAAAT	cohl_al2.seq
13442	GCTCGAATGTTCTGATCTAAGCGTTCAATATTGCGGTTAAAGGTGACAAAT	a909_al2.seq
	ACCCGCTAAATATTTTCATGTTCTATGCTCTTTTCTAAAAATCTCTAAATAA	Majority
	13560 13570 13580 13590 13600	
13539	ACCCGCTAAATATTTTCATGTTCTATGCTCTTTTCTAAAAATCTCTAAATAA	cohl_al2.seq
13492	ACCCGCTAAATATTTTCATGTTCTATGCTCTTTTCTAAAAATCTCTAAATAA	a909_al2.seq
	CTGAATGACTGGTGGTTTGGTTATAAAAAACGATACCGACATAGATAGTTA	Majority
	13610 13620 13630 13640 13650	
13589	CTGAATGACTGGTGGTTTGGTTATAAAAAACGATACCGACATAGATAGTTA	cohl_al2.seq
13542	CTGAATGACTGGTGGTTTGGTTATAAAAAACGATACCGACATAGATAGTTA	a909_al2.seq

FIGURE 21T

131/487

	CTGCTTACCTAAACCTTTGAATGACATAATTTACCAATGATACTGACATTTTGA	Majority
	13660 13670 13680 13690 13700	
13639	CTGCTACTAAACCTTTGAATGACATAATTTACCAATGATACTGACATTTTGA	coh1_al2.seq
13592	CTGCTACTAAACCTTTGAATGACATAATTTACCAATGATACTGACATTTTGA	a909_al2.seq
	GTATTGATATAATAGAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA	Majority
	13710 13720 13730 13740 13750	
13689	GTATTGATATAATAGAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA	coh1_al2.seq
13642	GTATTGATATAATAGAGTACAGCTCCACTAAGAGTAGCAGCAATTAAATA	a909_al2.seq
	GCCGAGCATTTCCTCTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG	Majority
	13760 13770 13780 13790 13800	
13739	GCCGAGCATTTCCTCTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG	coh1_al2.seq
13692	GCCGAGCATTTCCTCTTGTTAATTCTTTTAAAGTAAATACATCTCTTAAAG	a909_al2.seq
	AGATAGCTTTGATATAGGGAGACAATAAATTCAGTAATAAAGTGTAGAGATA	Majority
	13810 13820 13830 13840 13850	
13789	AGATAGCTTTGATATAGGGAGACAATAAATTCAGTAATAAAGTGTAGAGATA	coh1_al2.seq
13742	AGATAGCTTTGATATAGGGAGACAATAAATTCAGTAATAAAGTGTAGAGATA	a909_al2.seq
	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC	Majority
	13860 13870 13880 13890 13900	
13839	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC	coh1_al2.seq
13792	ATAGCTCCCATAGCACCTAAATTTGGTATTAAAAAGTATATTAAGCACAAAC	a909_al2.seq
	ATTTGCCACAAGTCCAATAAAGTGCAGACATTGTGTAAGCTTTTGTACGTC	Majority
	13910 13920 13930 13940 13950	
13889	ATTTGCCACAAGTCCAATAAAGTGCAGACATTGTGTAAGCTTTTGTACGTC	coh1_al2.seq
13842	ATTTGCCACAAGTCCAATAAAGTGCAGACATTGTGTAAGCTTTTGTACGTC	a909_al2.seq
	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA	Majority
	13960 13970 13980 13990 14000	
13939	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA	coh1_al2.seq
13892	TTGAAGCCAGTAGATACTGTGTCCCTAAAGCGTTACCATAAGAAATGCAA	a909_al2.seq
	ATGATCATCAAA	Majority
	14010	
13989	ATGATCATCAAA	coh1_al2.seq
13942	ATGATCATCAAA	a909_al2.seq

Decoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 21U

132/487

M K L S K K L L F S A A V L T M V A G S T V E P V A Q F A T G M S I V R A A E V S Q E R P A K T T V Majority																																																			
10										20										30										40										50											
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_2603.pep
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_a909.pep
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_cjb111.pep
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_cohl.pep
1	M	K	L	S	K	K	L	L	F	S	A	A	V	L	T	M	V	A	G	S	T	V	E	P	V	A	Q	F	A	T	G	M	S	I	V	R	A	A	E	V	S	Q	E	R	P	A	K	T	T	V	sag645_nem316.pep
N I Y K L Q A D S Y K S E I T S N G G I E N K D G E V I S N Y A K L G D N V K G L Q C V Q F K R Y K Majority																																																			
60										70										80										90										100											
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_2603.pep
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_a909.pep
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_cjb111.pep
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_cohl.pep
51	N	I	Y	K	L	Q	A	D	S	Y	K	S	E	I	T	S	N	G	G	I	E	N	K	D	G	E	V	I	S	N	Y	A	K	L	G	D	N	V	K	G	L	Q	C	V	Q	F	K	R	Y	K	sag645_nem316.pep
V K T D I S V D E L K K L T T V E A A D A K V G T I L E E G V S L P Q K T N A Q C L V D A L D S K Majority																																																			
110										120										130										140										150											
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_2603.pep	
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_a909.pep	
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_cjb111.pep	
01	V	K	T	D	I	S	V	D	E	L	K	K	L	T	T	V	E	A	A	D	A	K	V	G	T	I	L	E	E	G	V	S	L	P	Q	K	T	N	A	Q	C	L	V	D	A	L	D	S	K	sag645_nem316.pep	
S N V R Y L Y V E D L K N S P S N I T K A Y A V P F V L E L P V A N S T G T G F L S E I N I Y P K N Majority																																																			
160										170										180										190										200											
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_2603.pep
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_a909.pep
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_cjb111.pep
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_cohl.pep
51	S	N	V	R	Y	L	Y	V	E	D	L	K	N	S	P	S	N	I	T	K	A	Y	A	V	P	F	V	L	E	L	P	V	A	N	S	T	G	T	G	F	L	S	E	I	N	I	Y	P	K	N	sag645_nem316.pep
V V T D E P K T D K D V K K L G Q D D A G Y T I G E E F K W F L K S T I P A N L G D Y E K F E I T D Majority																																																			
210										220										230										240										250											
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_2603.pep
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_a909.pep
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_cjb111.pep
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_cohl.pep
01	V	V	T	D	E	P	K	T	D	K	D	V	K	K	L	G	Q	D	D	A	G	Y	T	I	G	E	E	F	K	W	F	L	K	S	T	I	P	A	N	L	G	D	Y	E	K	F	E	I	T	D	sag645_nem316.pep
K F A D G L T Y K S V G K I K I G S K T L N R D E H Y T I D E P T V D N Q N T L K I T F K P E K F K Majority																																																			
260										270										280										290										300											
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_2603.pep
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_a909.pep
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_cjb111.pep
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_cohl.pep
51	K	F	A	D	G	L	T	Y	K	S	V	G	K	I	K	I	G	S	K	T	L	N	R	D	E	H	Y	T	I	D	E	P	T	V	D	N	Q	N	T	L	K	I	T	F	K	P	E	K	F	K	sag645_nem316.pep
E I A E L L K G M T L V K N Q D A L D K A T A N T D D A A F L E I P V A S T I N E K A V L G K A I E Majority																																																			
310										320										330										340										350											
1	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_2603.pep
1	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_a909.pep
1	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_cjb111.pep
1	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_cohl.pep
1	E	I	A	E	L	L	K	G	M	T	L	V	K	N	Q	D	A	L	D	K	A	T	A	N	T	D	D	A	A	F	L	E	I	P	V	A	S	T	I	N	E	K	A	V	L	G	K	A	I	E	sag645_nem316.pep
N T F E L Q Y D H T P D K A D N P K P S N P P R K P E V H T G G K R F V K K D S T E T Q T L G G A E Majority																																																			
360										370										380										390										400											
1	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_2603.pep
1	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_a909.pep
1	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_cjb111.pep
1	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P	S	N	P	P	R	K	P	E	V	H	T	G	G	K	R	F	V	K	K	D	S	T	E	T	Q	T	L	G	G	A	E	sag645_cohl.pep
1	N	T	F	E	L	Q	Y	D	H	T	P	D	K	A	D	N	P	K	P																																

133/487

FDLLASDGTAVKWTDALEIKANTNKNYIACEAVTCQPIKLSHTDGTFEIK Majority									
410		420		430		440		450	
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
01	F	D	L	L	A	S	D	G	T
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTD Majority									
460		470		480		490		500	
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
51	G	L	A	Y	A	V	D	A	N
ITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGA AVMAFAVKGMKRR Majority									
510		520		530		540		550	
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
01	I	T	V	D	S	A	D	A	T
TKDN Majority									
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					
51	T	K	D	N					

scoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 22A

134/487

M K K R Q K I W R G L S V T L L I L S Q I P F G I L V Q G E T Q D T N Q A L G K V I V K K T G D N A Majority																																																				
10				20				30				40				50																																				
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_2603.pep	
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_coh1.pep	
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_cjb111.pep	
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_18rs21.pep	
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_nem316.pep	
1	M	K	K	R	Q	K	I	W	R	G	L	S	V	T	L	L	I	L	S	Q	I	P	F	G	I	L	V	Q	G	E	T	Q	D	T	N	Q	A	L	G	K	V	I	V	K	K	T	G	D	N	A	sag649_cjb111.pep	
T P L G K A T F V L K N D N D K S E T S H E T V E G S G E A T F F E N I K P G D Y T L R E E T A P I G Majority																																																				
60				70				80				90				100																																				
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_2603.pep
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_coh1.pep
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_cjb111.pep
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_18rs21.pep
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_nem316.pep
51	T	P	L	G	K	A	T	F	V	L	K	N	D	N	D	K	S	E	T	S	H	E	T	V	E	G	S	G	E	A	T	F	F	E	N	I	K	P	G	D	Y	T	L	R	E	E	T	A	P	I	G	sag649_cjb111.pep
Y K K T D K T W K V K V A D N G A T I I E G M D A D K A E K R K E V L N A Q Y P K S A I Y E D T K E Majority																																																				
110				120				130				140				150																																				
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_2603.pep	
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_coh1.pep	
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_cjb111.pep	
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_18rs21.pep	
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_nem316.pep	
101	Y	K	K	T	D	K	T	W	K	V	K	V	A	D	N	G	A	T	I	I	E	G	M	D	A	D	K	A	E	K	R	K	E	V	L	N	A	Q	Y	P	K	S	A	I	Y	E	D	T	K	E	sag649_cjb111.pep	
N Y P L V N V E G S K V G E Q Y K A L N P I N G K D G R R E I A E G W L S K K I T G V N D L D K N K Majority																																																				
160				170				180				190				200																																				
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_2603.pep	
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_coh1.pep	
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_cjb111.pep	
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_18rs21.pep	
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_nem316.pep	
151	N	Y	P	L	V	N	V	E	G	S	K	V	G	E	Q	Y	K	A	L	N	P	I	N	G	K	D	G	R	R	E	I	A	E	G	W	L	S	K	K	I	T	G	V	N	D	L	D	K	N	K	sag649_cjb111.pep	
Y K I E L T V E G K T T V E T K E L N Q P L D V V V L L D N S N S M N N E R A N N S Q R A L K A G E Majority																																																				
210				220				230				240				250																																				
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_2603.pep	
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_coh1.pep	
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_cjb111.pep	
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_18rs21.pep	
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_nem316.pep	
201	Y	K	I	E	L	T	V	E	G	K	T	T	V	E	T	K	E	L	N	Q	P	L	D	V	V	V	L	L	D	N	S	N	S	M	N	N	E	R	A	N	N	S	Q	R	A	L	K	A	G	E	sag649_cjb111.pep	
A V E K L I D K I T S N K D N R V A L V T Y A S T I F D G T E A T V S K G V A D Q N G K A L N D S V Majority																																																				
260				270				280				290				300																																				
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_2603.pep	
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_coh1.pep	
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_cjb111.pep	
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_18rs21.pep	
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_nem316.pep	
251	A	V	E	K	L	I	D	K	I	T	S	N	K	D	N	R	V	A	L	V	T	Y	A	S	T	I	F	D	G	T	E	A	T	V	S	K	G	V	A	D	Q	N	G	K	A	L	N	D	S	V	sag649_cjb111.pep	
S W D Y H K T T F T A T T H N Y S Y L N L T N D A N E V N I L K S R I P K E A E H I N G D R T L Y Q Majority																																																				
310				320				330				340				350																																				
301	S	W	D	Y	H	K	T	T	F	T	A	T	T	H	N	Y	S	Y	L	N	L	T	N	D	A	N	E	V	N	I	L	K	S	R	I	P	K	E	A	E	H	I	N	G	D	R	T	L	Y	Q	sag649_2603.pep	
301	S	W	D	Y	H	K	T	T	F	T	A	T	T	H	N	Y	S	Y	L	N	L	T	N	D	A	N	E	V	N	I	L	K	S	R	I	P	K	E	A	E	H	I	N	G	D	R	T	L	Y	Q	sag649_coh1.pep	
301	S	W	D	Y	H	K	T	T	F	T	A	T	T	H	N	Y	S	Y	L	N	L	T	N	D	A	N	E	V	N	I	L	K	S	R	I	P	K	E	A	E	H	I	N	G	D	R	T	L	Y	Q	sag649_cjb111.pep	
301	S	W	D	Y	H	K	T	T	F	T	A	T	T	H	N	Y	S	Y	L	N	L	T	N	D	A	N	E	V	N	I	L	K	S	R	I	P	K	E	A	E	H	I	N	G	D	R	T	L	Y	Q	sag649_18rs21.pep	
301	S	W	D	Y	H	K	T	T	F	T	A	T	T	H	N	Y	S	Y	L	N	L	T	N	D	A	N	E	V	N	I	L	K	S	R	I	P	K	E	A	E	H	I	N	G	D	R	T	L				

Figure 23

135/487

T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V Majority									
410		420		430		440		450	
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_2603.pep							
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_cohl.pep							
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_cjb111.pep							
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_18rs21.pep							
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_nem316.pep							
401	T S Y Q N Q F N S F L N K I P D R S G I L Q E D F I I N G D D Y Q I V K G D G E S F K L F S D R K V	sag649_cjb111.pep							
P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T Majority									
460		470		480		490		500	
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_2603.pep							
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_cohl.pep							
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_cjb111.pep							
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_18rs21.pep							
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_nem316.pep							
451	P V T G C T T Q A A Y R V P Q N Q L S V M S N E G Y A I N S G Y I Y L Y W R D Y N W V Y P F D P K T	sag649_cjb111.pep							
K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E Majority									
510		520		530		540		550	
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_2603.pep							
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_cohl.pep							
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_cjb111.pep							
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_18rs21.pep							
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_nem316.pep							
501	K K V S A T K Q I K T H G E P T T L Y F N G N I R P K G Y D I F T V G I G V N G D P G A T P L E A E	sag649_cjb111.pep							
K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G Majority									
560		570		580		590		600	
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_2603.pep							
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_cohl.pep							
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_cjb111.pep							
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_18rs21.pep							
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_nem316.pep							
551	K F M Q S I S S K T E N Y T N V D D T N K I Y D E L N K Y F K T I V E E K H S I V D G N V T D P M G	sag649_cjb111.pep							
E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T Majority									
610		620		630		640		650	
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_2603.pep							
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_cohl.pep							
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_cjb111.pep							
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_18rs21.pep							
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_nem316.pep							
601	E M I E F Q L K N G Q S F T H D D Y V L V G N D G S Q L K N G V A L G G P N S D G G I L K D V T V T	sag649_cjb111.pep							
Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K Majority									
660		670		680		690		700	
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_2603.pep							
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_cohl.pep							
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_cjb111.pep							
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_18rs21.pep							
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_nem316.pep							
651	Y D K T S Q T I K I N H L N L G S G Q K V V L T Y D V R L K D N Y I S N K F Y N T N N R T T L S P K	sag649_cjb111.pep							
S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L Majority									
710		720		730		740		750	
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_2603.pep							
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_cohl.pep							
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_cjb111.pep							
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_18rs21.pep							
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_nem316.pep							
701	S E K E P N T I R D F P I P K I R D V R E F P V L T I S N Q K K M G E V E F I K V N K D K H S E S L	sag649_cjb111.pep							
L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S Majority									
760		770		780		790		800	
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_2603.pep							
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_cohl.pep							
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_cjb111.pep							
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_18rs21.pep							
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_nem316.pep							
751	L G A K F Q L Q I E K D F S G Y K Q F V P E G S D V T T K N D G K I Y F K A L Q D G N Y K L Y E I S	sag649_cjb111.pep							

FIGURE 23A

136/487

	810	820	830	840	850	Majority
	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
801	SPDGYIEVKT	TKPVVTF	TIQNGEV	TNLKADPN	ANKKNQIGYLE	CNGKHLITN
	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
	860	870	880	890		
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS
851	TPKRPPG	VFPK	TGGIG	TI	VYILV	GSTFMILTICS

Majority

sag649_2603.pep
sag649_coh1.pep
sag649_cjb111.pep
sag649_18rs21.pep
sag649_nem316.pep
sag649_cjb111.pep

Decoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 23B

Figure 24

138/487

SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF Majority									
410 420 430 440 450									
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_2603.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_515.pep
401	SYAINNFKLGCASYESQFEQMKKNGYLNKSNFLLTDKPEDIKNGGESYFLF								sag1408_18rs21.seq
396	SYA I N S F Y K G S T Y A N Q F E R K F E R G Y L D K A N Y F I T D D P E K I K N G G E S Y F L F								sag1408_cjb111.pep
401	SYA I N N F K L G A S Y E S Q F E Q M K K N G Y L N K S N F L L T D K P E D I K N G G E S Y F L F								sag1408_h36b.pep
PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS Majority									
460 470 480 490 500									
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_2603.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_515.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_18rs21.seq
446	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_cjb111.pep
451	PLDSYQTQIIISGNLQKLHYLDLNLNYPKGTIYRNGPVEHGTPTKLYINS								sag1408_h36b.pep
LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT Majority									
510 520 530 540 550									
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_2603.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_515.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_18rs21.seq
496	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_cjb111.pep
501	LKQKNYDIFNFGIDISGFRQVYNEFYKKNQDGTFFQKLKEEAFELSDGEIT								sag1408_h36b.pep
ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI Majority									
560 570 580 590 600									
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_2603.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_515.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_18rs21.seq
546	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_cjb111.pep
551	ELMRSFSSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKEINSIVNGTI								sag1408_h36b.pep
EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK Majority									
610 620 630 640 650									
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								sag1408_2603.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								sag1408_515.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								sag1408_18rs21.seq
596	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								sag1408_cjb111.pep
601	EDPMGDKINLQLGNGQTLQPSDYTLQGNDSVMKDGCIATGCPNNDGGILK								sag1408_h36b.pep
GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL Majority									
660 670 680 690 700									
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_2603.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_515.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_18rs21.seq
646	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_cjb111.pep
651	GVKLEYIGNKLYVRGLNLGEGQKVTLTYDVKLDDSFISNKFYDTNGRTTL								sag1408_h36b.pep
NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL Majority									
710 720 730 740 750									
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_2603.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_515.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_18rs21.seq
696	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_cjb111.pep
701	NPKSEDPNTLRDFFIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNKKL								sag1408_h36b.pep
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI Majority									
760 770 780 790 800									
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_2603.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_515.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_18rs21.seq
746	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_cjb111.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_h36b.pep
751	LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLI								sag1408_nem316.pep

FIGURE 24A

139/487

	810	820	830	840	850	Majority
	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_2603.pep
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_515.pep
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_18rs21.seq
796	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_cjb111.pep
801	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_h36b.pep
	E A V S P E D Y Q K I T N K P I L T F E V V K G S I Q N I I A V N K Q I S E Y H E E G D K H L I T N					sag1408_nen316.pep

	860	870	880	890	900	Majority
	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_2603.pep
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_515.pep
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_18rs21.seq
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_cjb111.pep
846	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_h36b.pep
851	T H I P P K G I I P M T G G K G I L S F I L I G G A M M S I A G G I Y I W K R Y K K S S D M S I E K					sag1408_nen316.pep

D

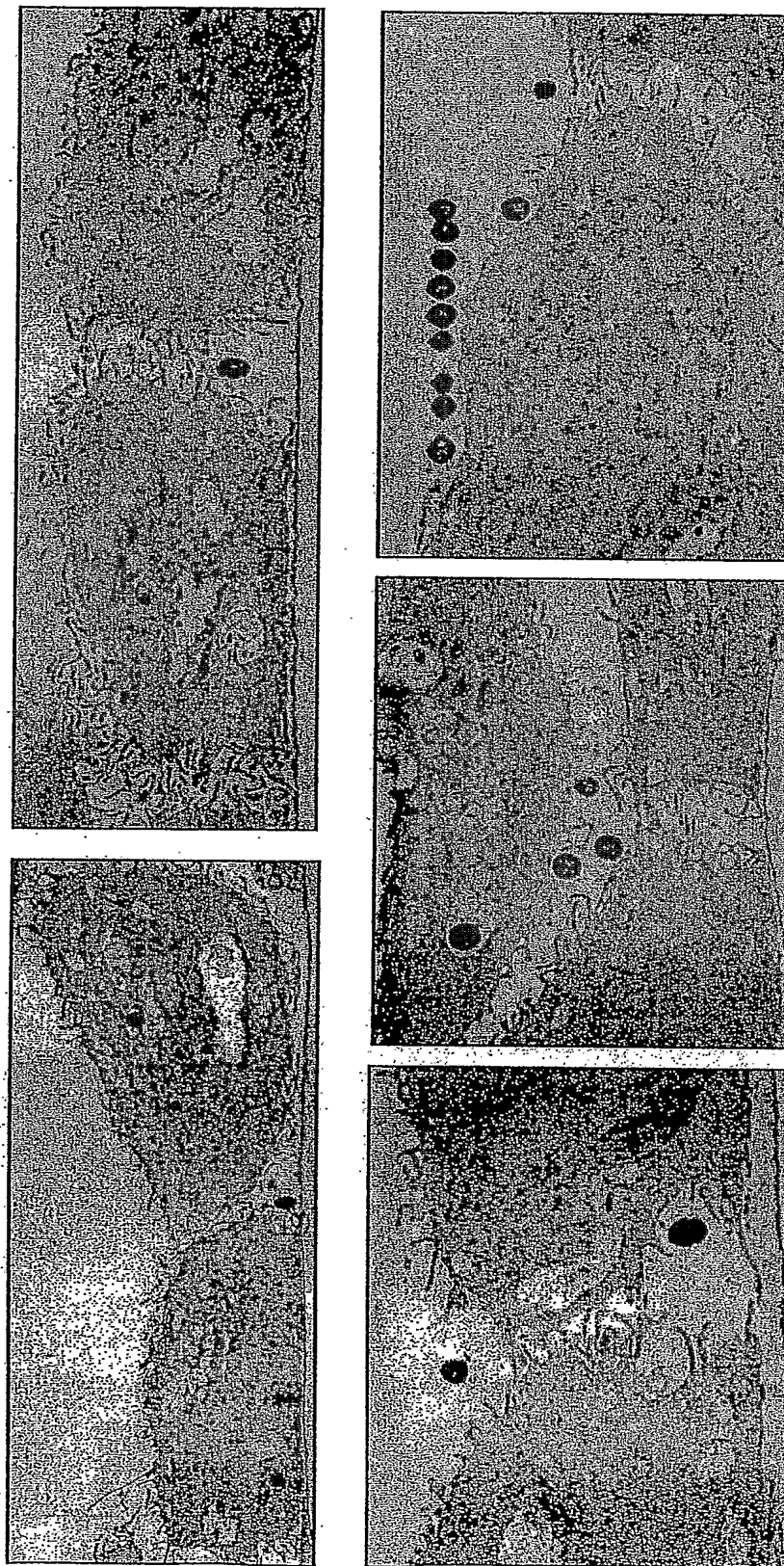
Majority

901 D	sag1408_2603.pep
901 D	sag1408_515.pep
901 D	sag1408_18rs21.seq
901 D	sag1408_cjb111.pep
896 D	sag1408_h36b.pep
901 D	sag1408_nen316.pep

Decoration 'Decoration #1': Shade (with solid black) residues that differ from the Consensus.

FIGURE 24B

Figure 25: GBS closely associate with tight junctions and cross the monolayer by a paracellular route



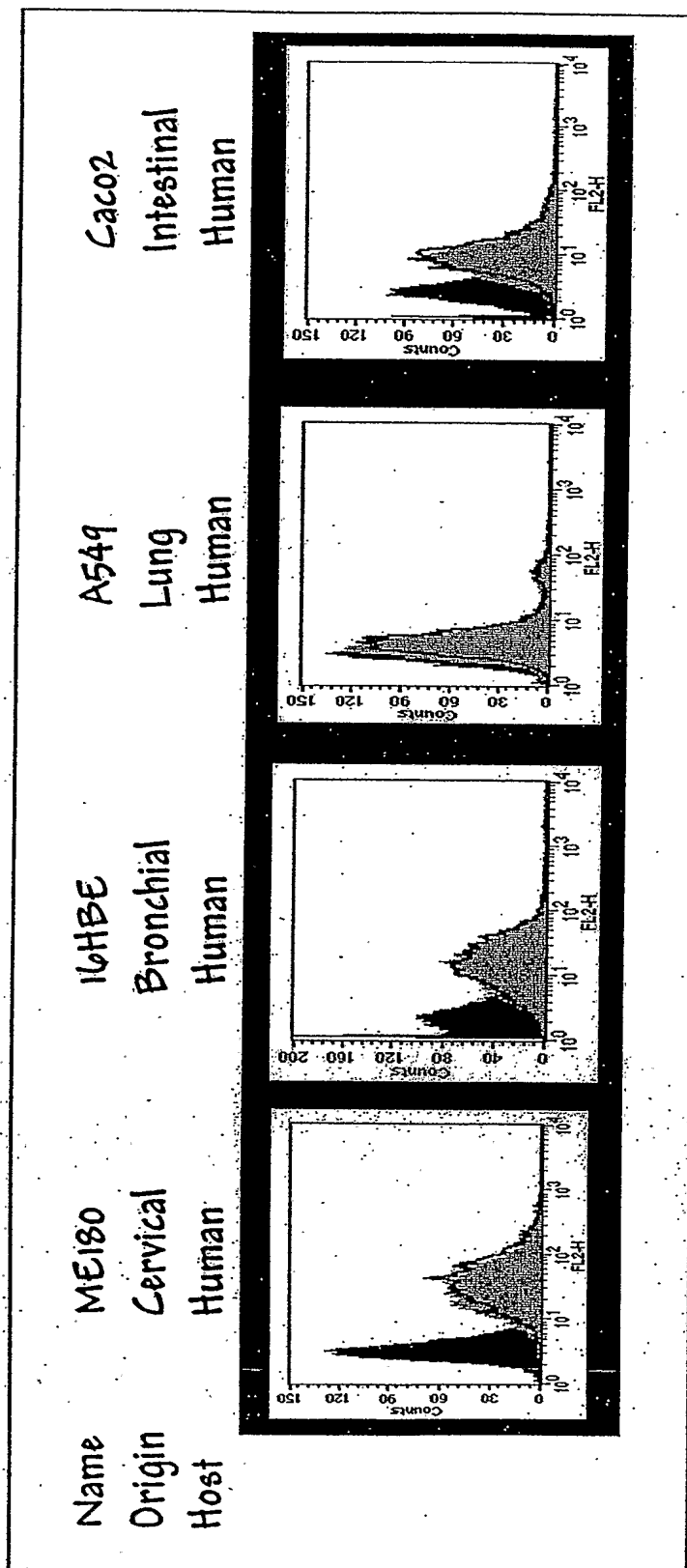
Transmission Electron Microscopy images of GBS infection of ME180 cervical epithelial cells.

Figure 26: GBS infection of ME180 cells
GBS infection of ME180 cells



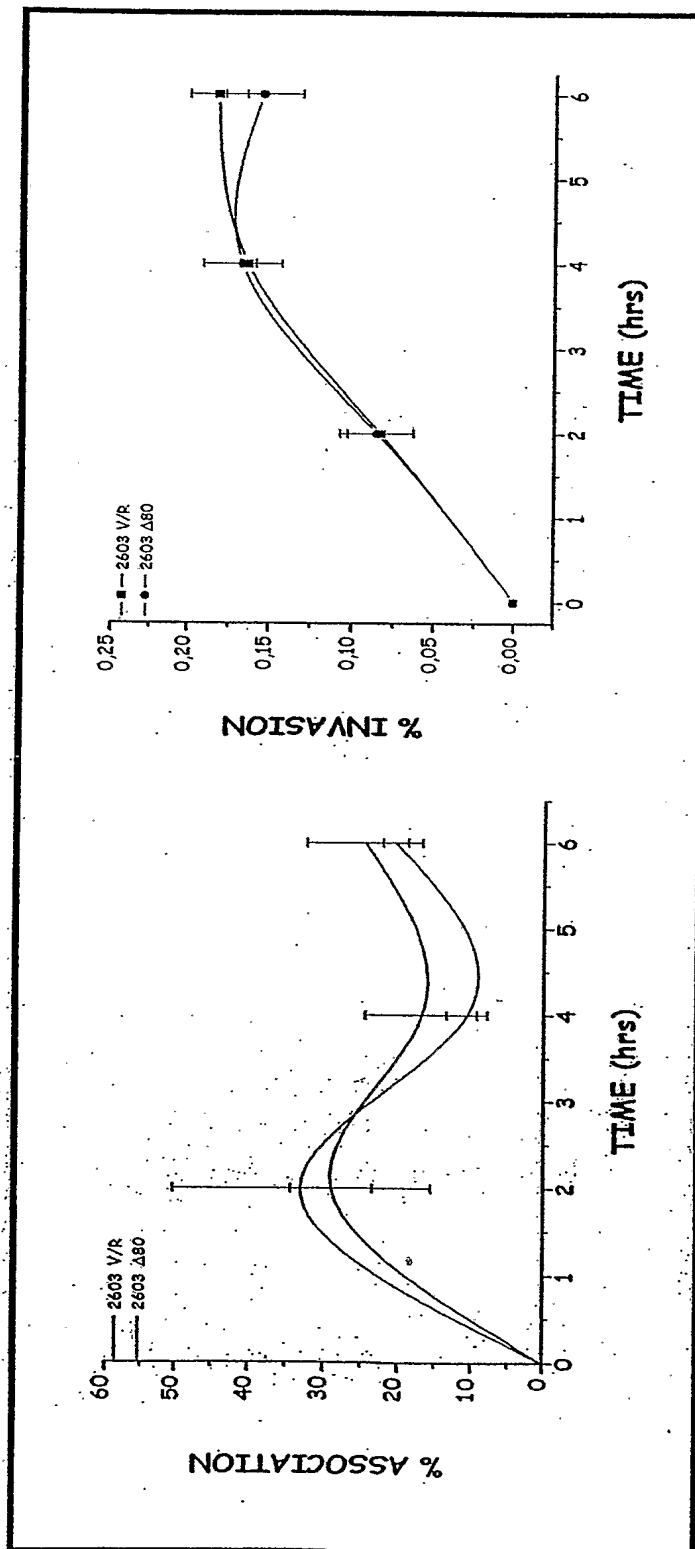
PCT/US05/27239/42/487

Figure 27



PCT/US05/27239 143/487

Figure 28



144/487

Figure 29

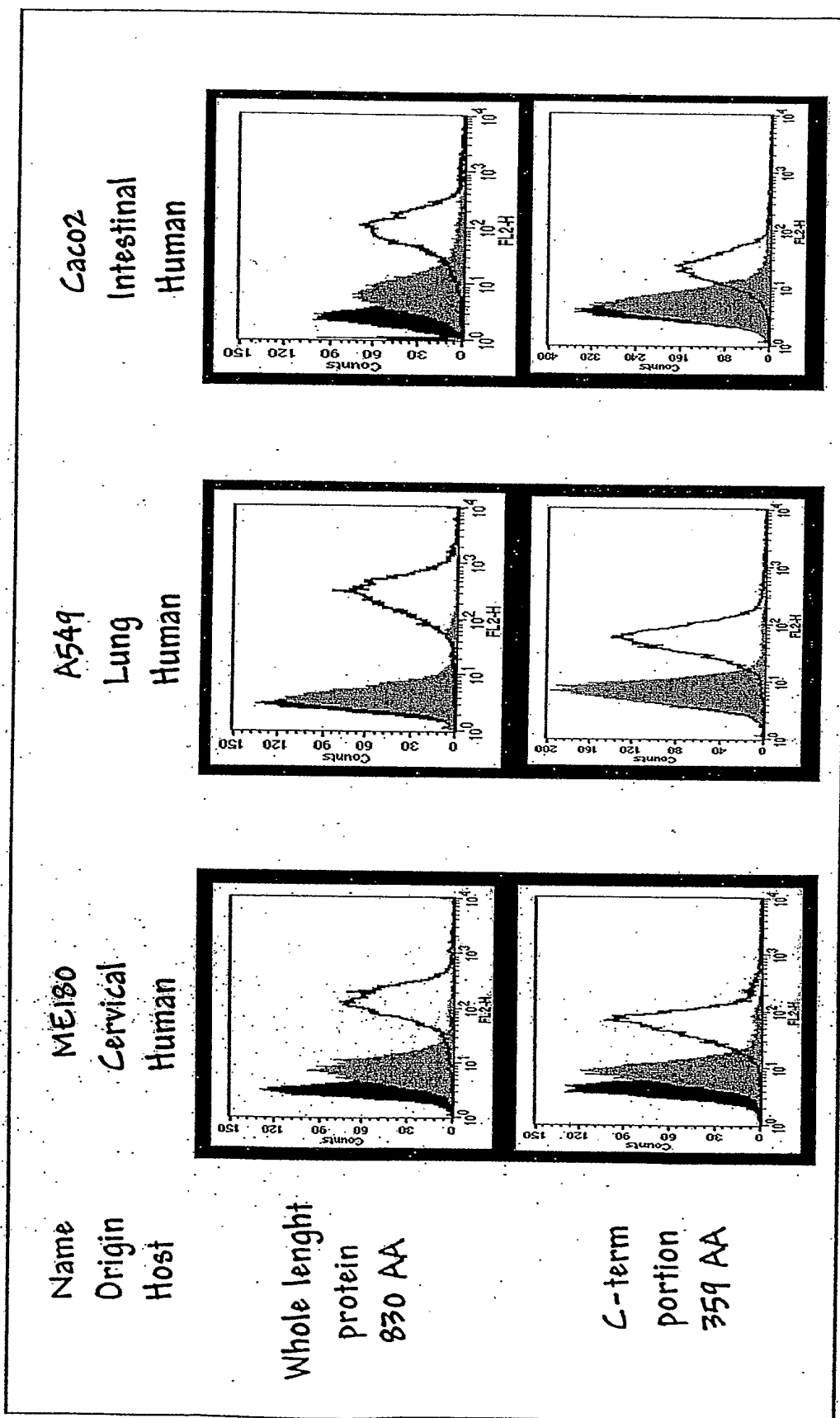
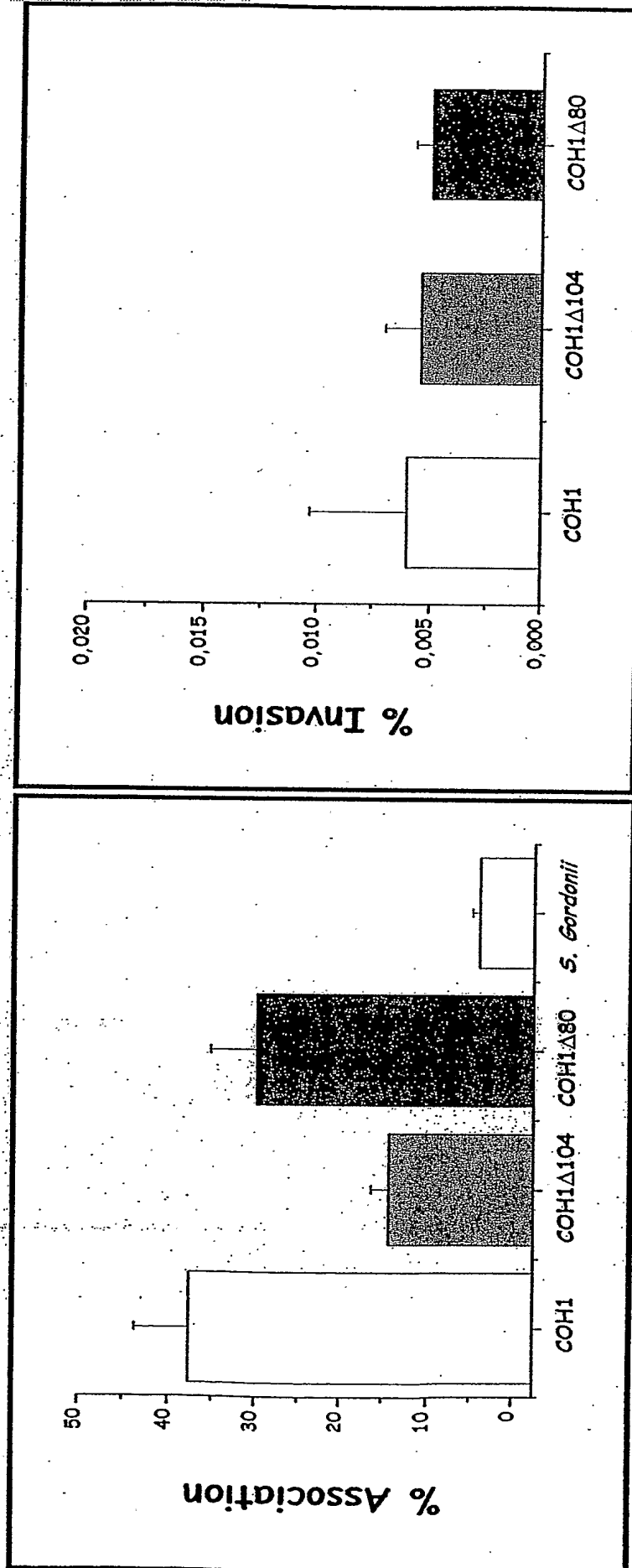
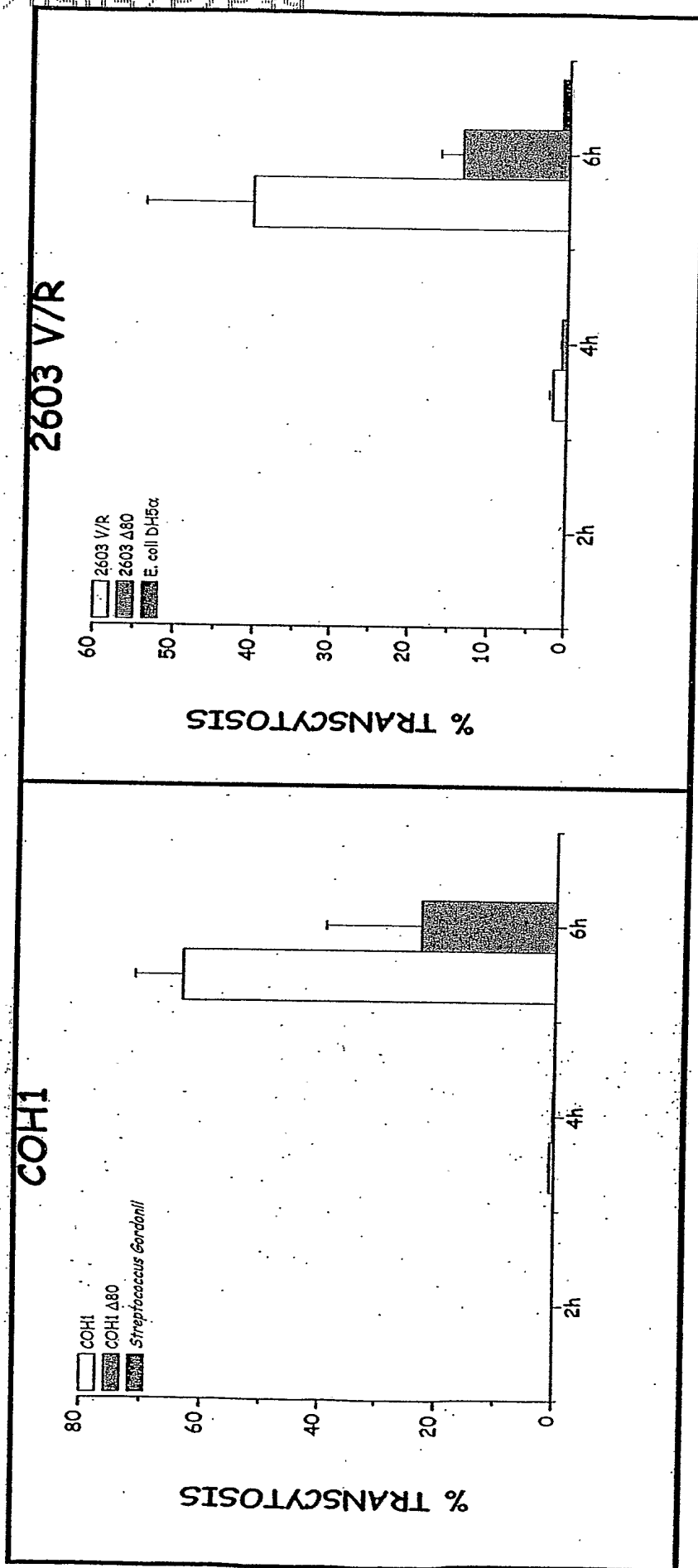


Figure 30



PCT/US2005/027239

Figure 31



PCT/US2005/027239/47/487

Figure 32

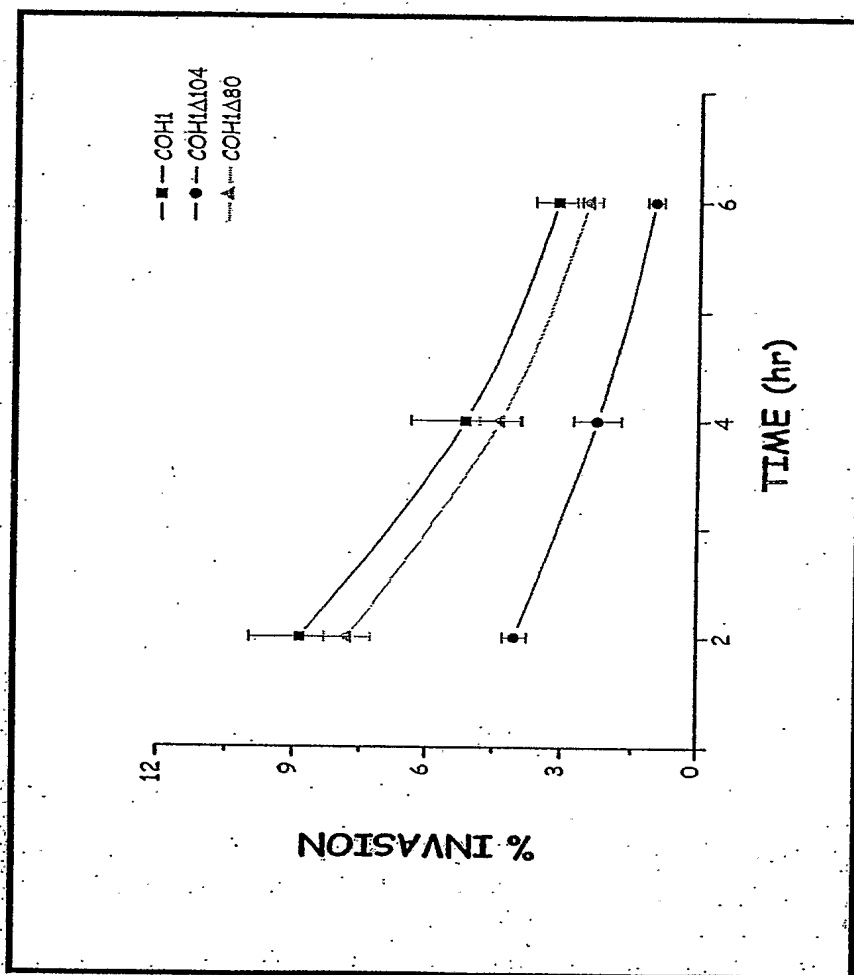
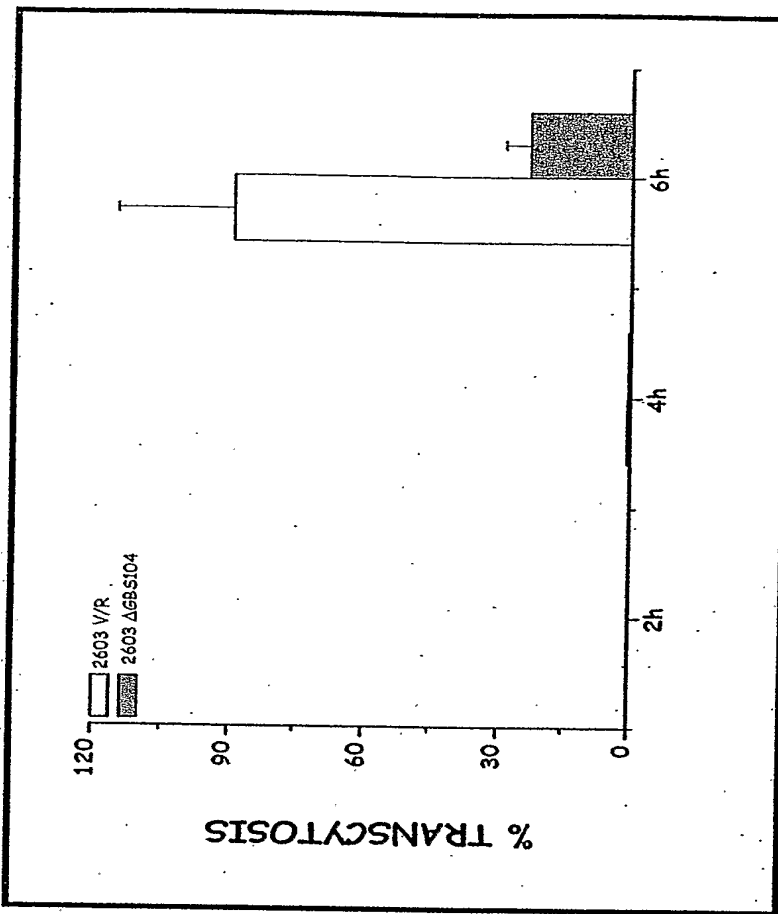
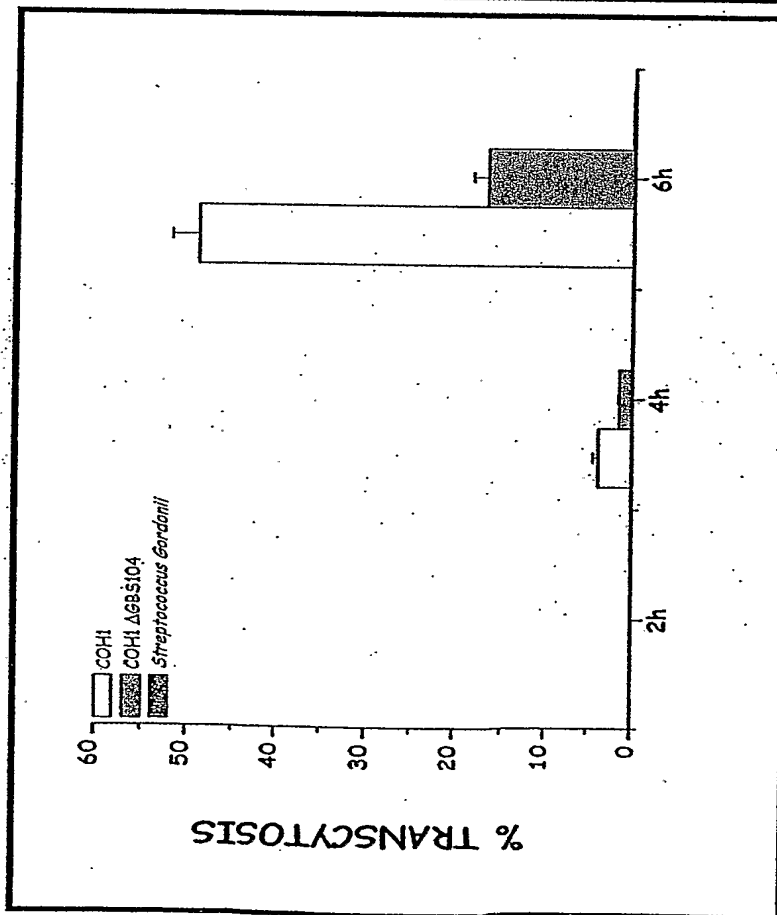


Figure 33

COH1

2603 V/R

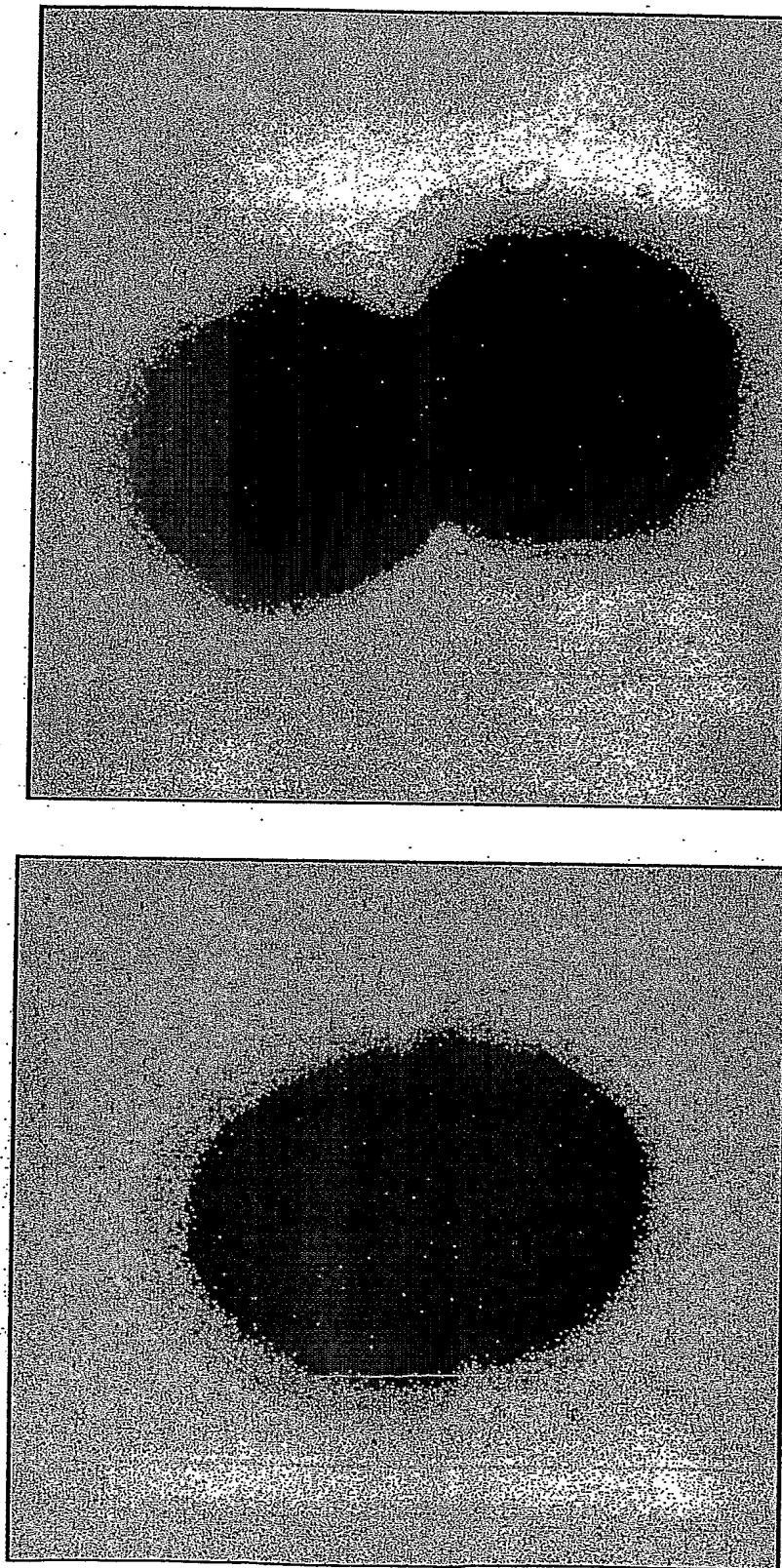


PCT/US05/27239 149/487

GBS STRAIN COH1 over GBS80

Figure 34

Negative staining EM



GBS STRAIN COH1 over GBS80

IEM anti-GBS80 (gold particles 10nm)

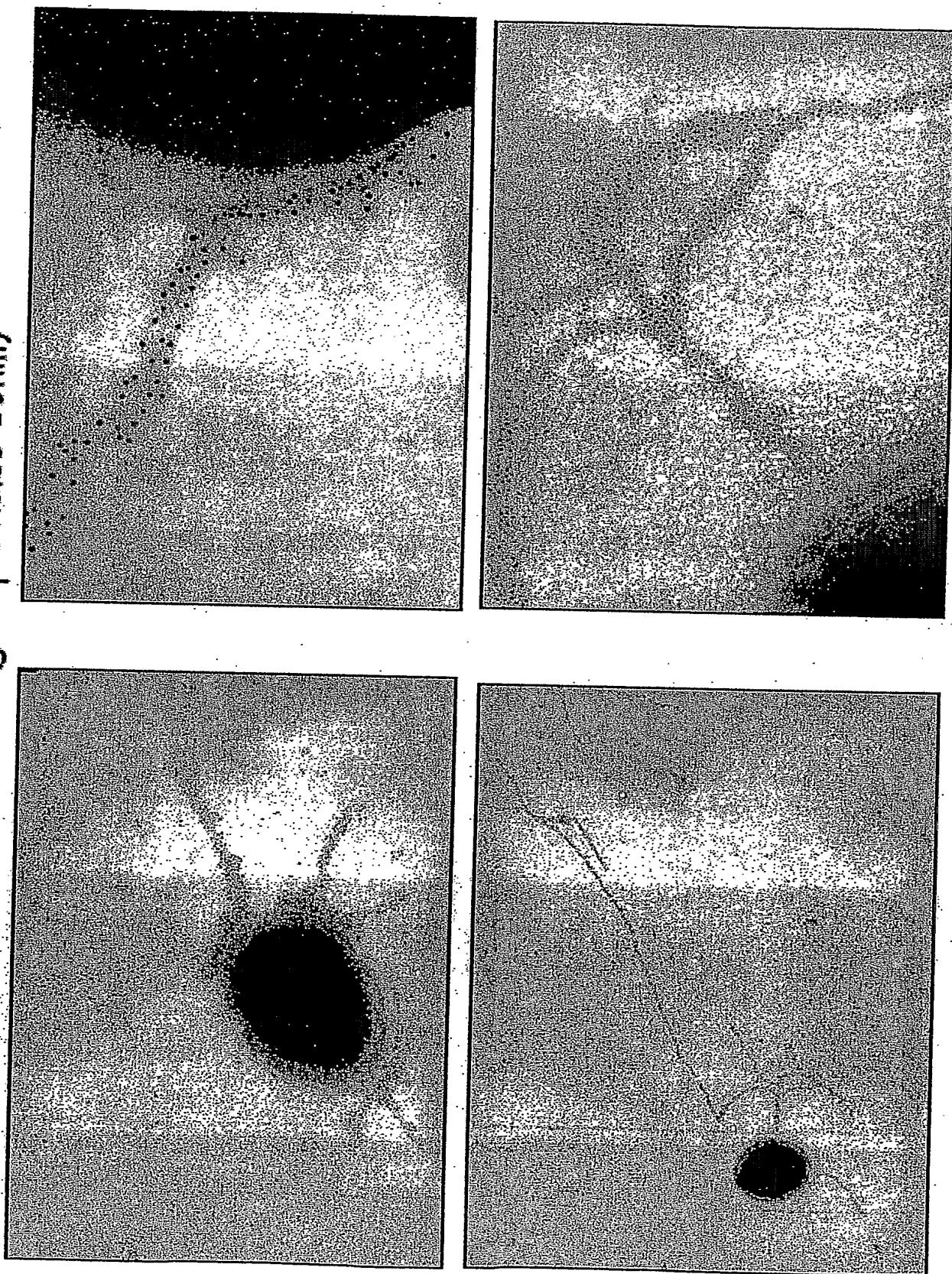
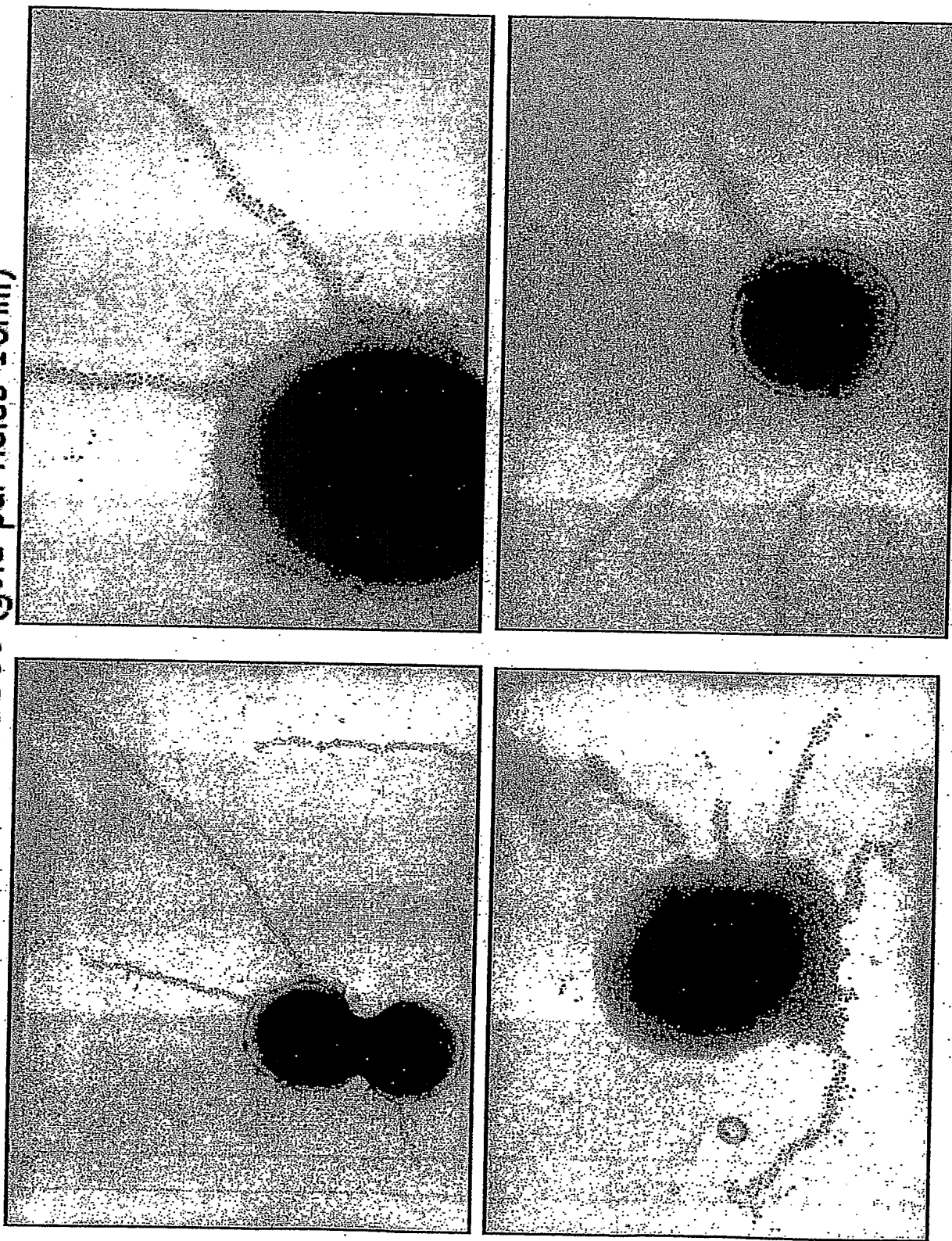


Figure 35

GBS STRAIN COH1 over GBS80

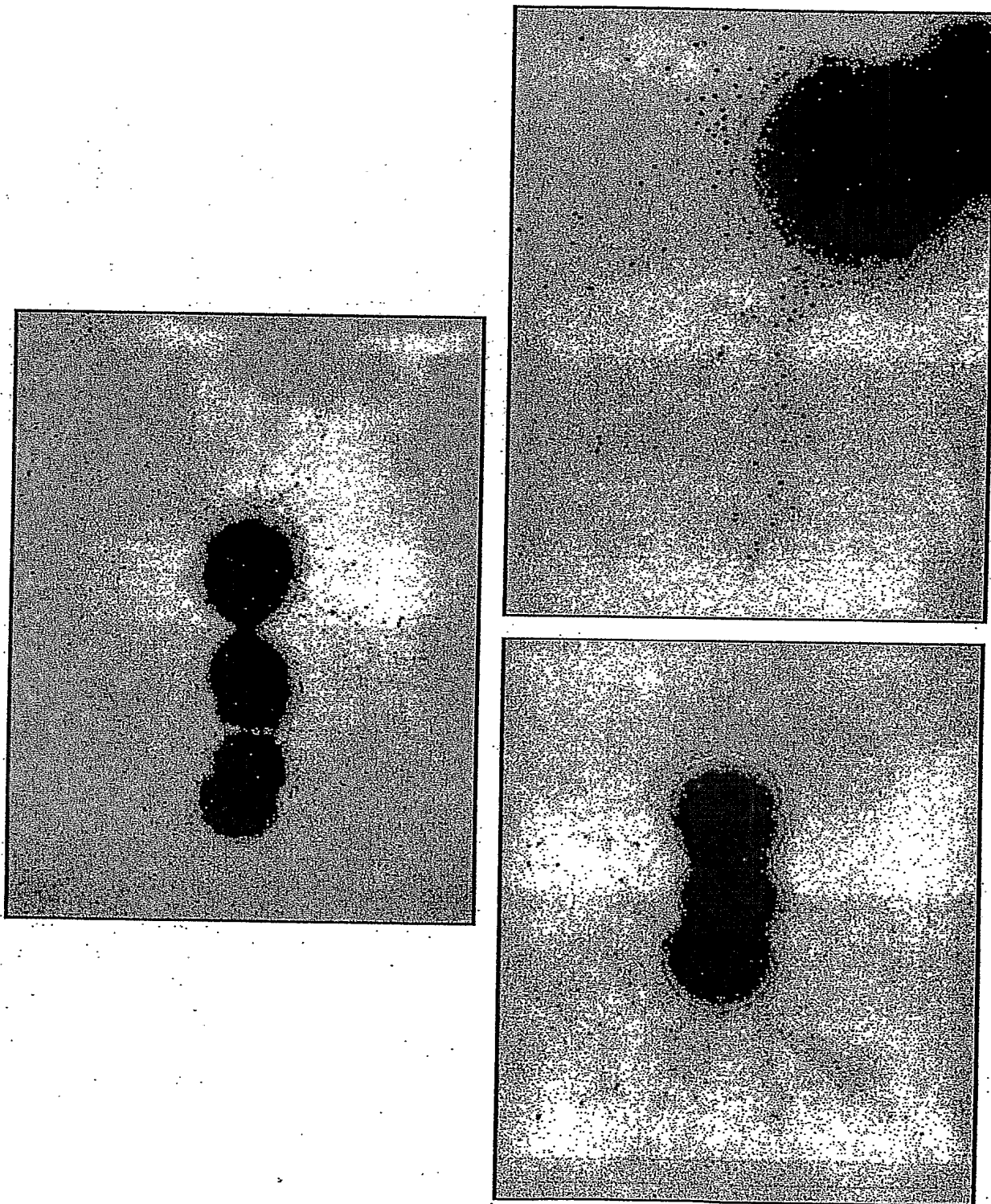
Figure 36

IEM anti-GBS80 (gold particles 10nm)



GBS STRAIN COH1 over GBS80
IEM anti-GBS80 (gold particles 20nm)

Figure 37



GBS STRAIN COH1 over GBS80

IEM anti-GBS104 (gold particles 10nm)

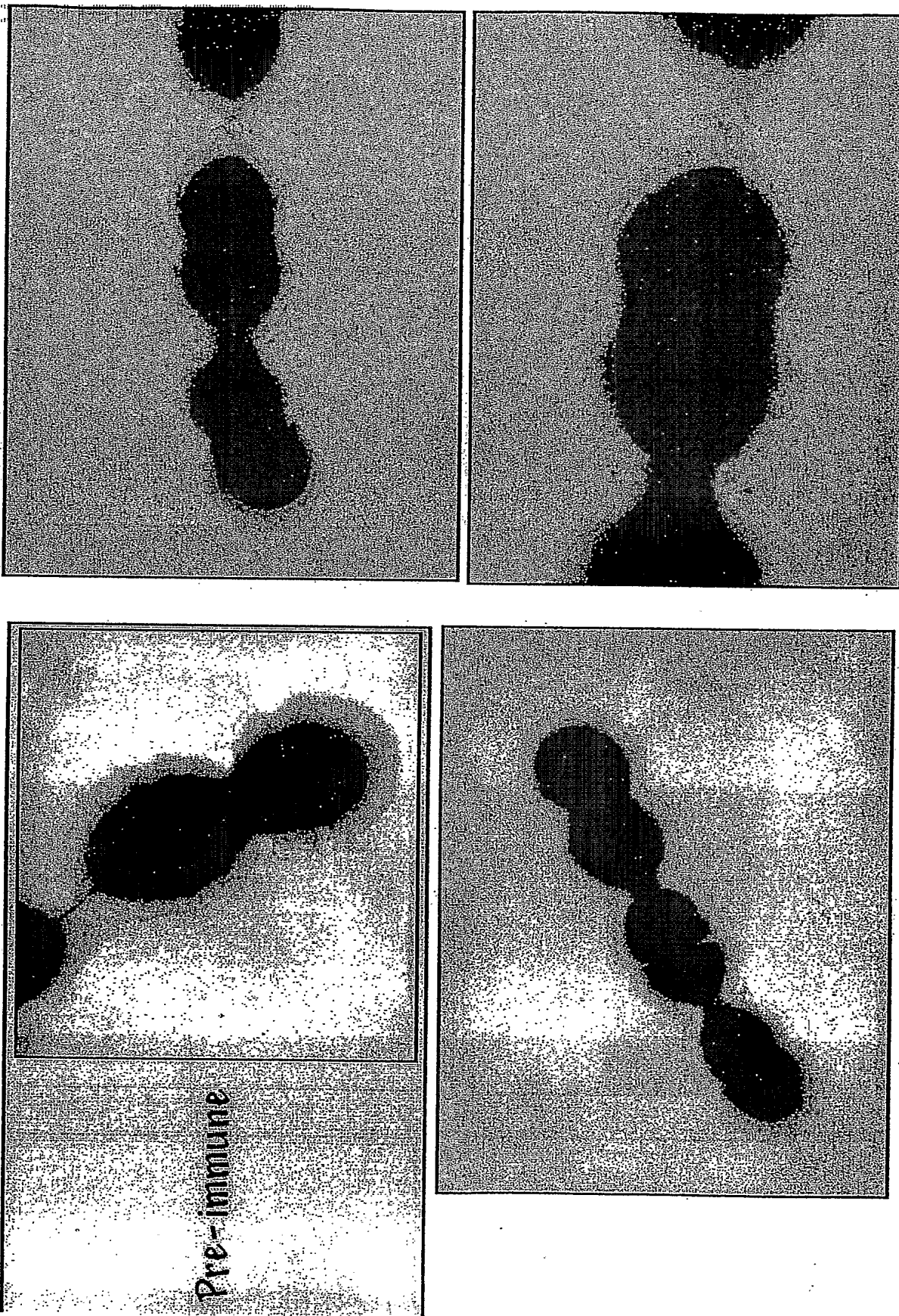


Figure 38

GBS STRAIN COH1 over GBS80

Figure 39

IEM anti-GBS80 (gold particles 20nm) anti-GBS104 (gold particles 10nm)

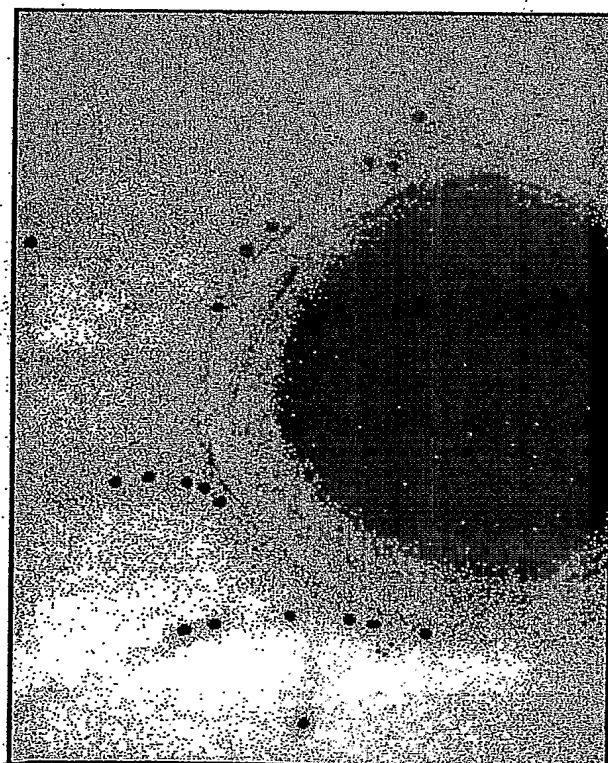
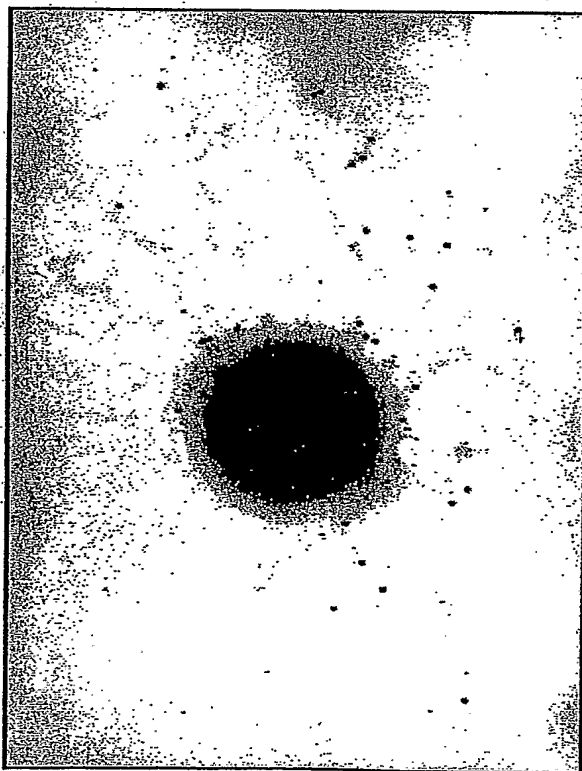
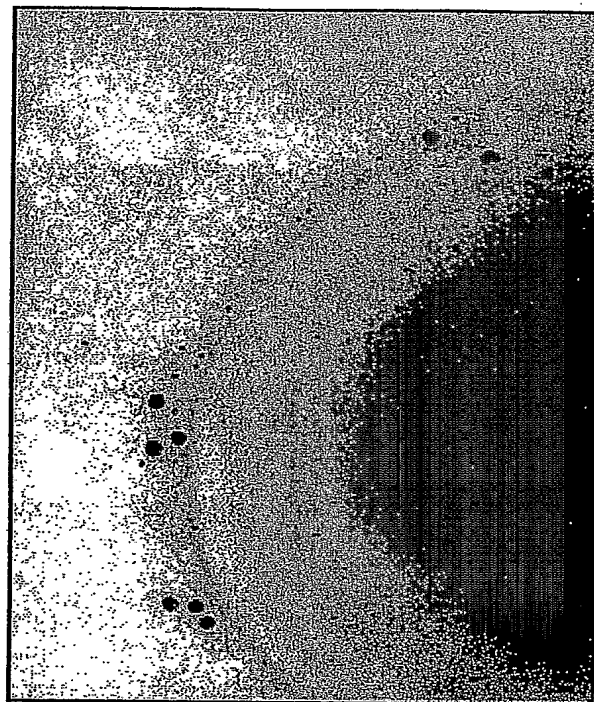
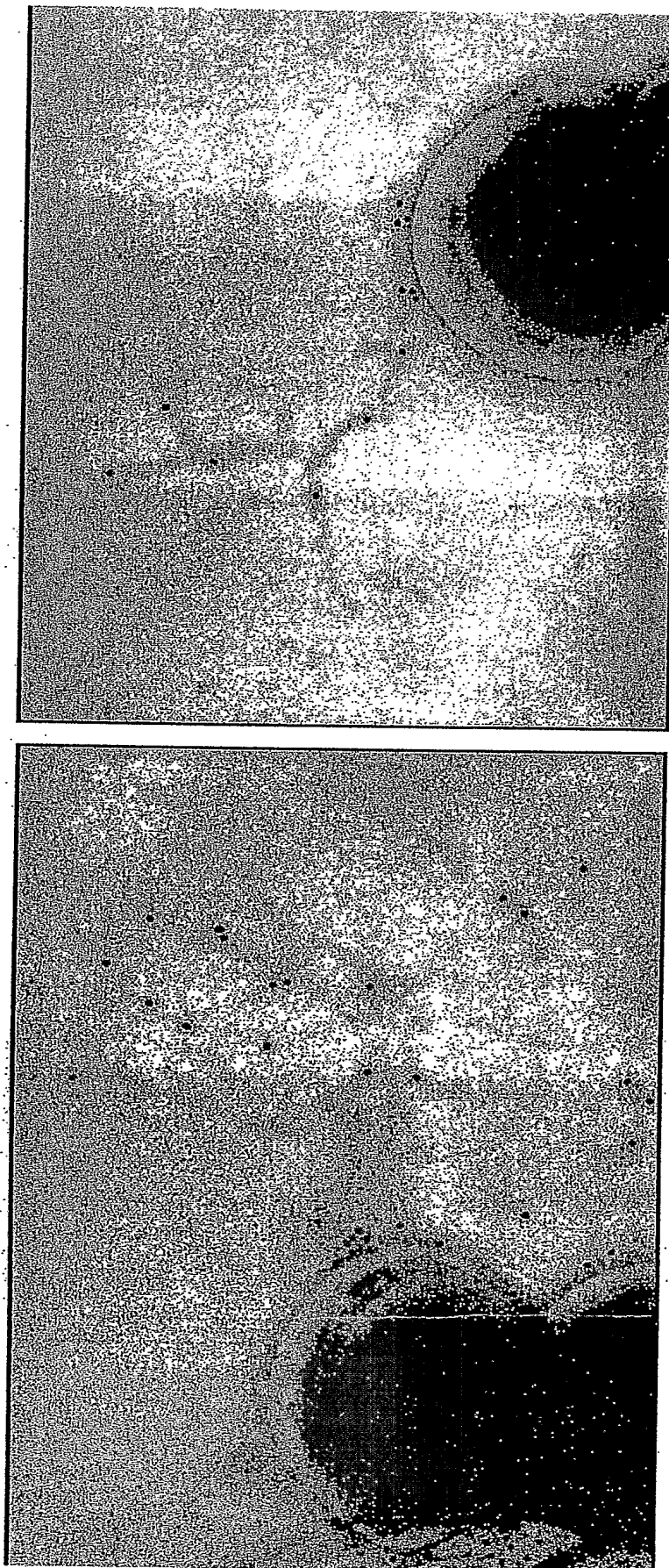


Figure 40

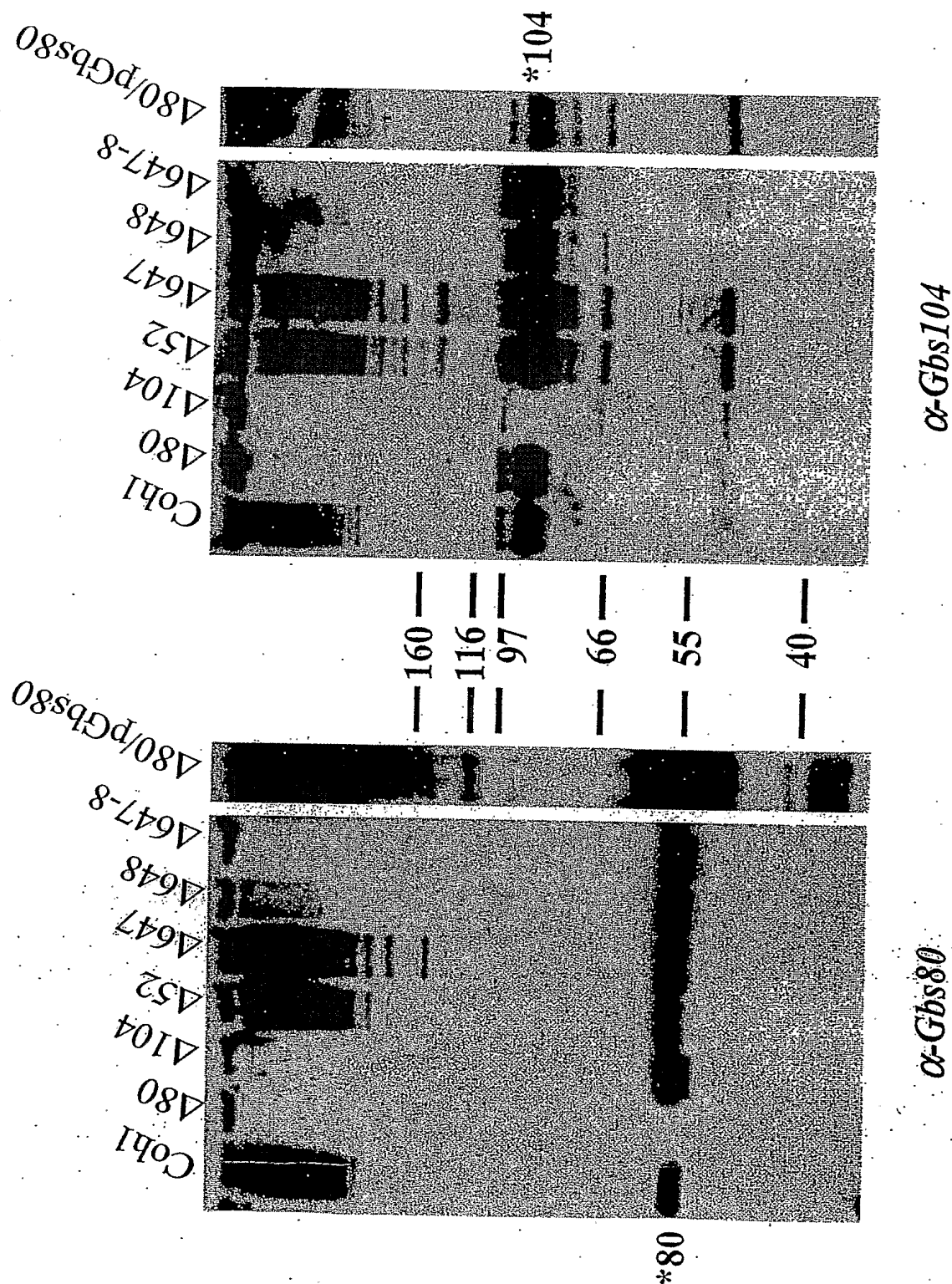
GBS STRAIN COH1 over GBS80

IEM anti-GBS80 (gold particles 20nm) anti-GBS104 (gold particles 10nm)



156/487

Figure 41: GBS 80 is necessary for polymer formation, GBS104 and sortase SAG0648 are necessary for efficient assembly



PCT/US2005/027239 157/487

Figure 42: Gbs67 is part of a second pilus;
Gbs80 is polymerized in strain 515
(515 lacks sortase 647-8, but has AI-2 sortases)

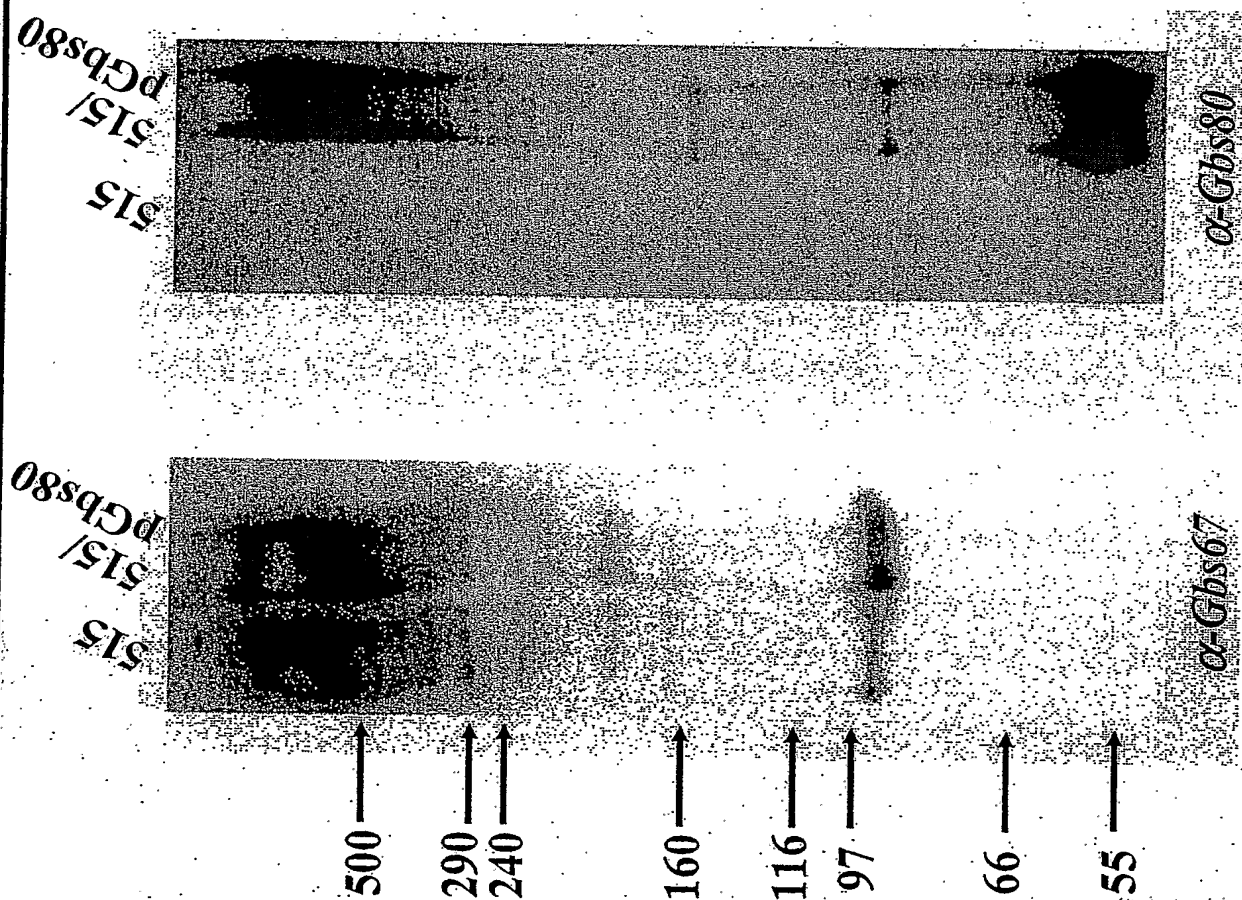
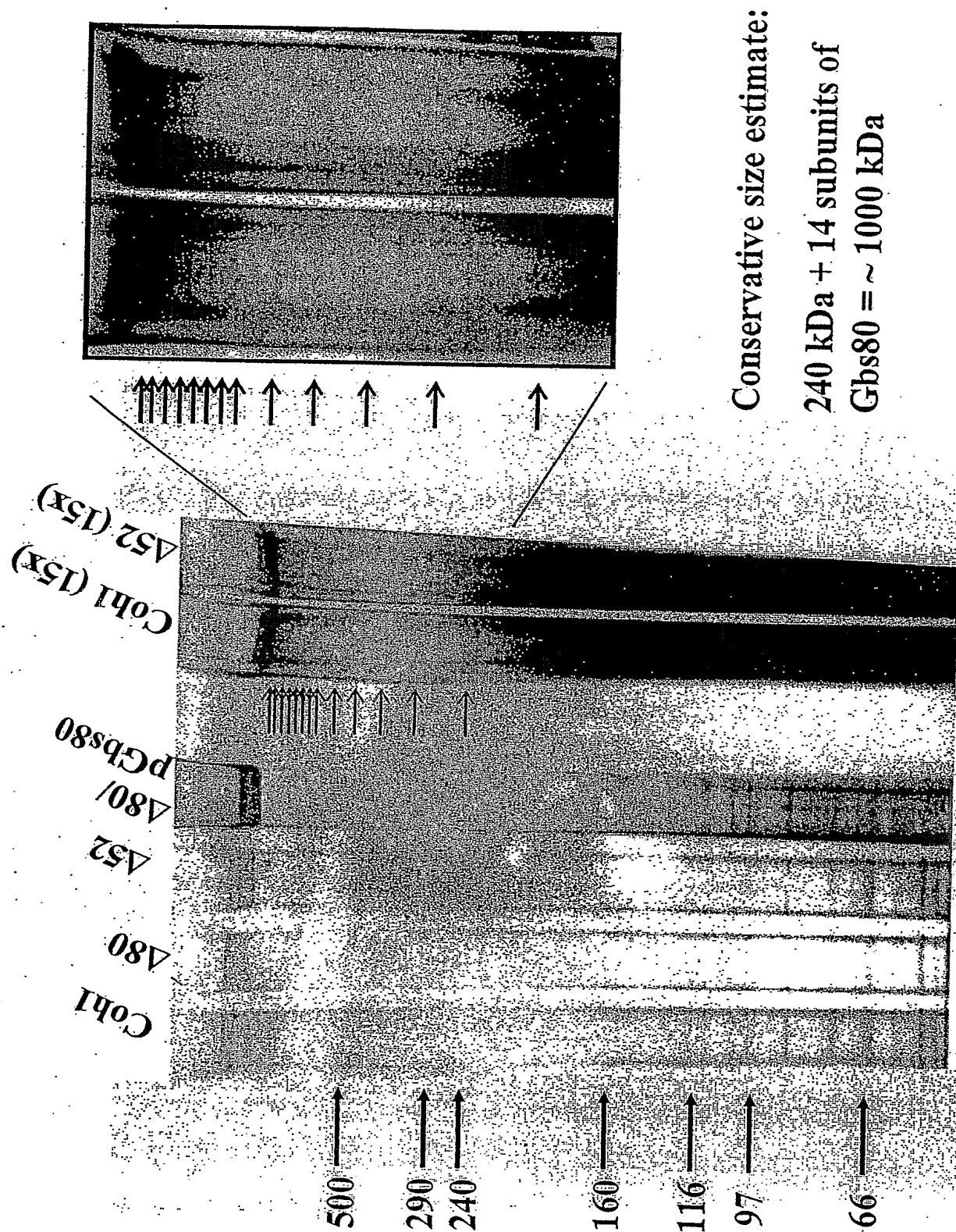


Figure 43: Two macro-molecules are visible in Coh1 at >1000 kDa, one is the Gbs80 pilin



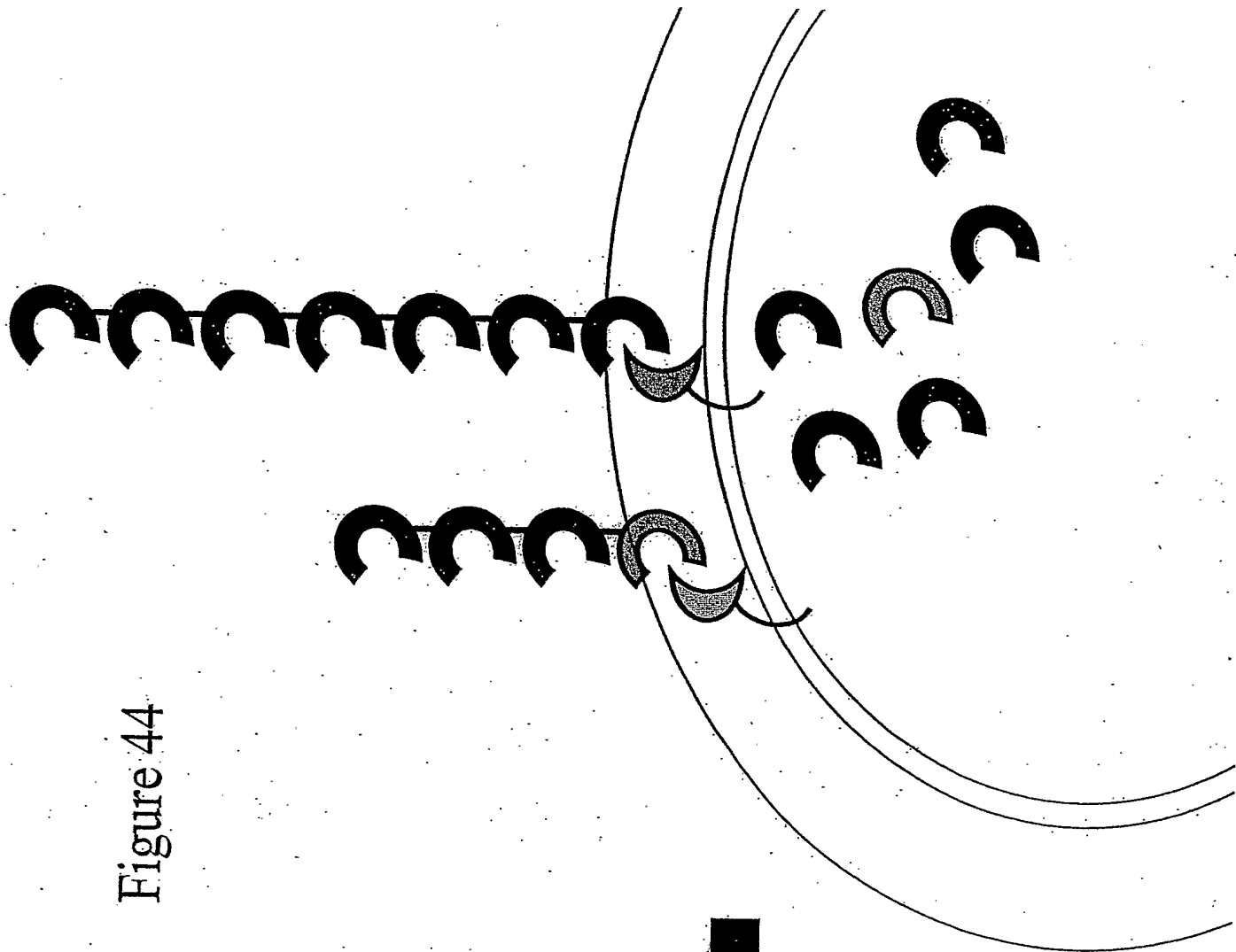


Figure 44

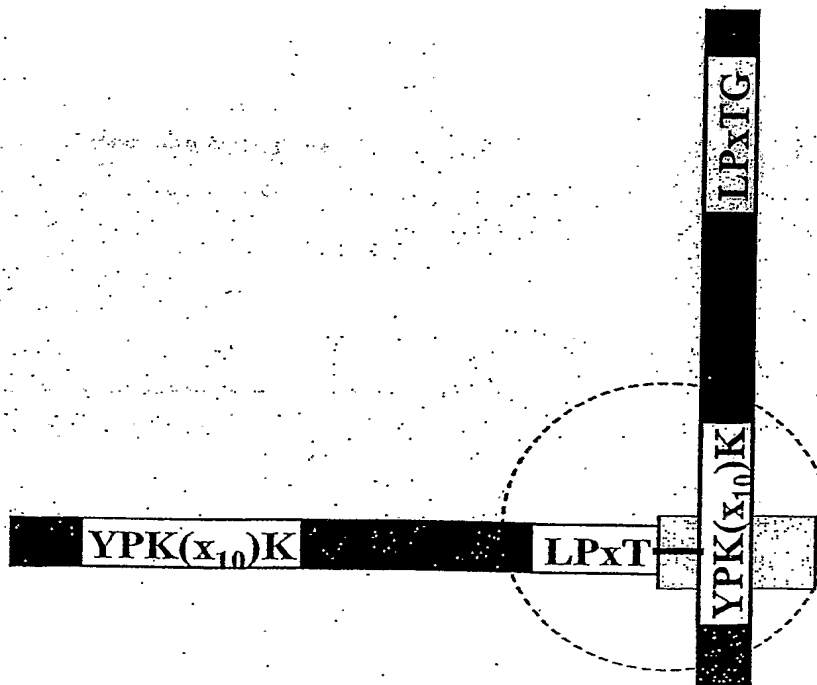


Figure 45: Gbs52 is a minor component of the GBS pilus

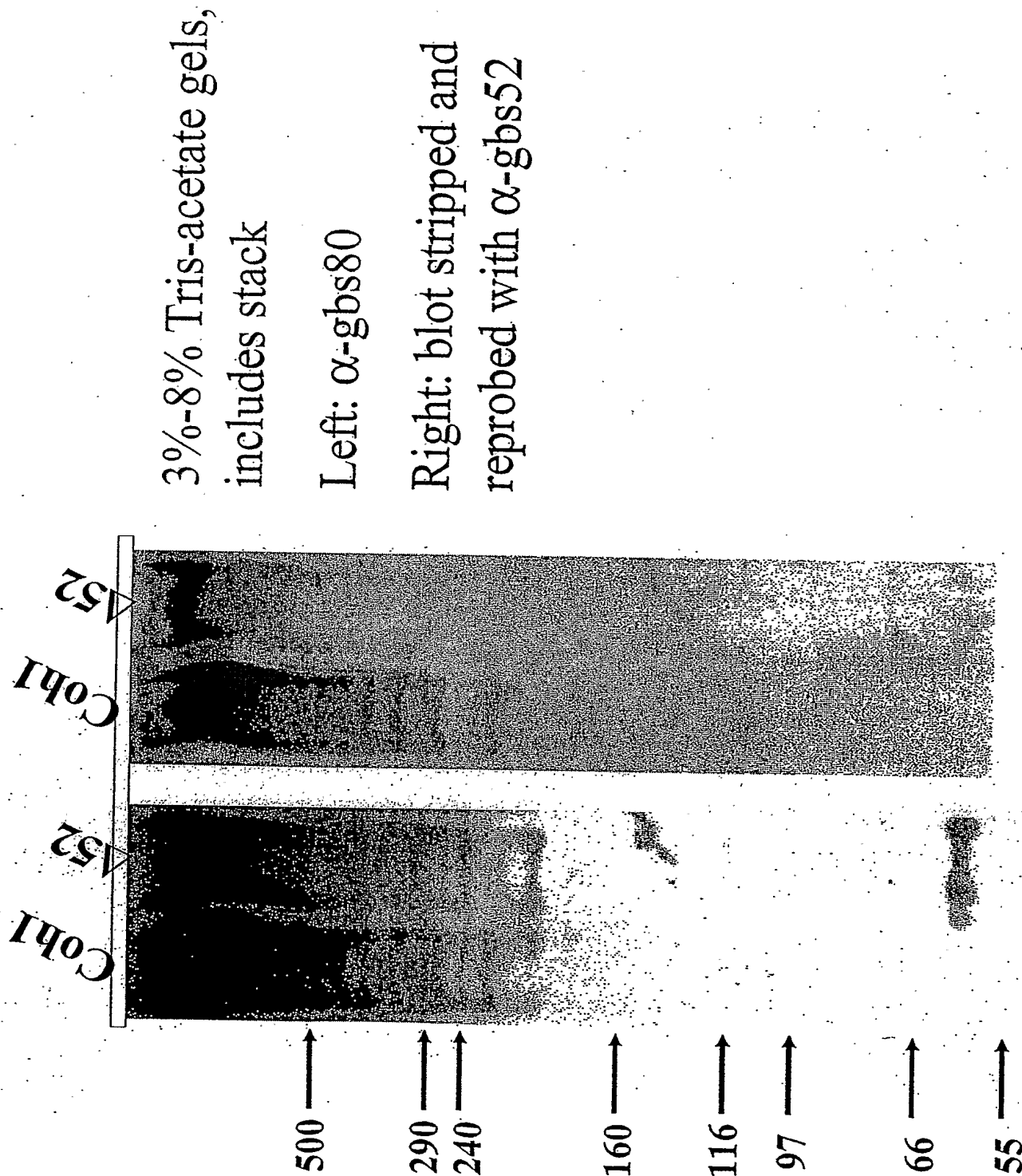


Figure 46: The pilus is found in the supernatant of the bacterial culture

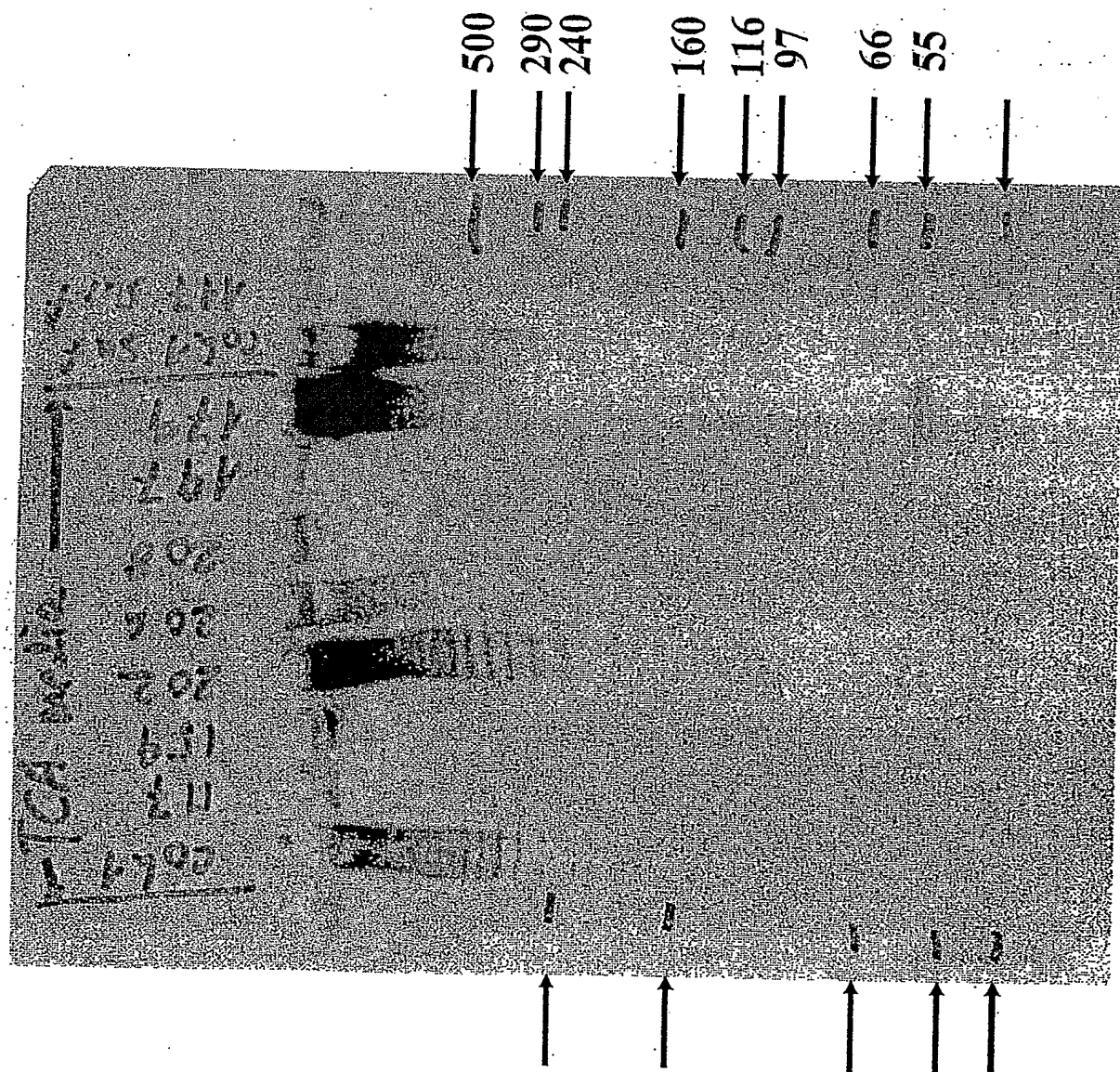
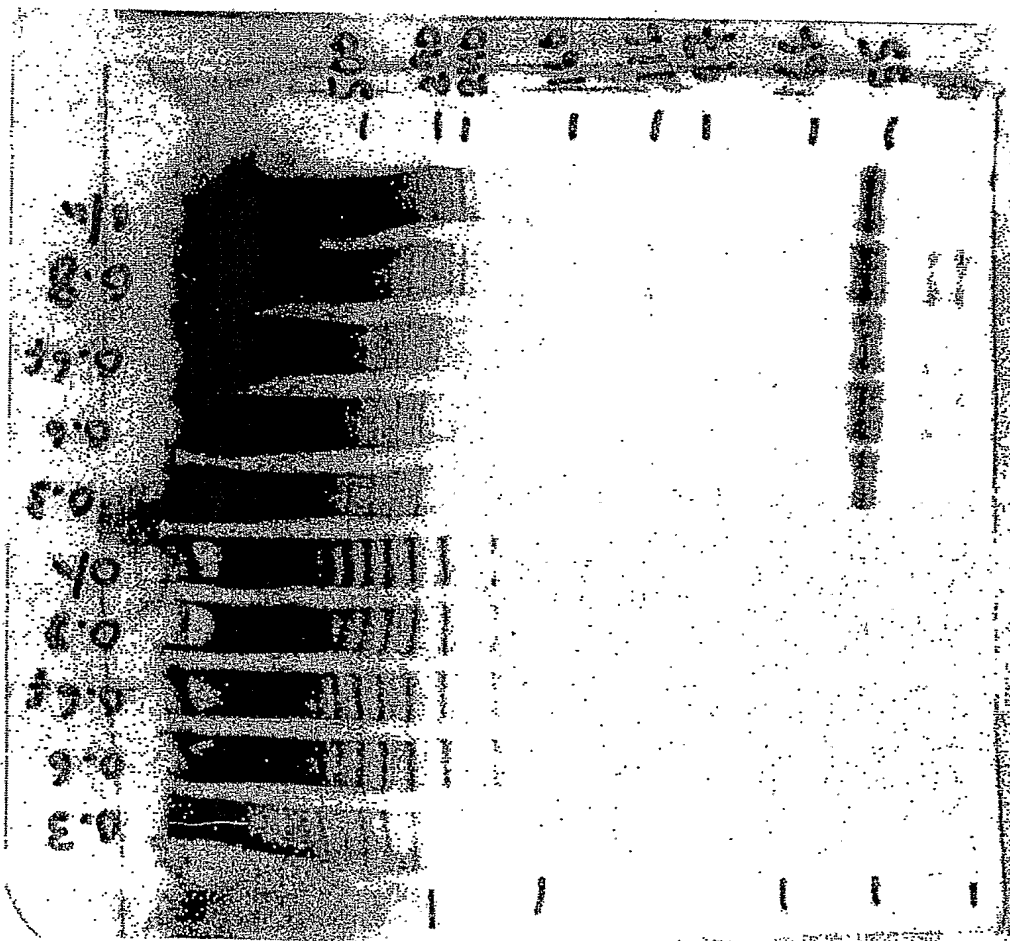


Figure 47: The pilus is found in the supernatant of cultures in all growth phases

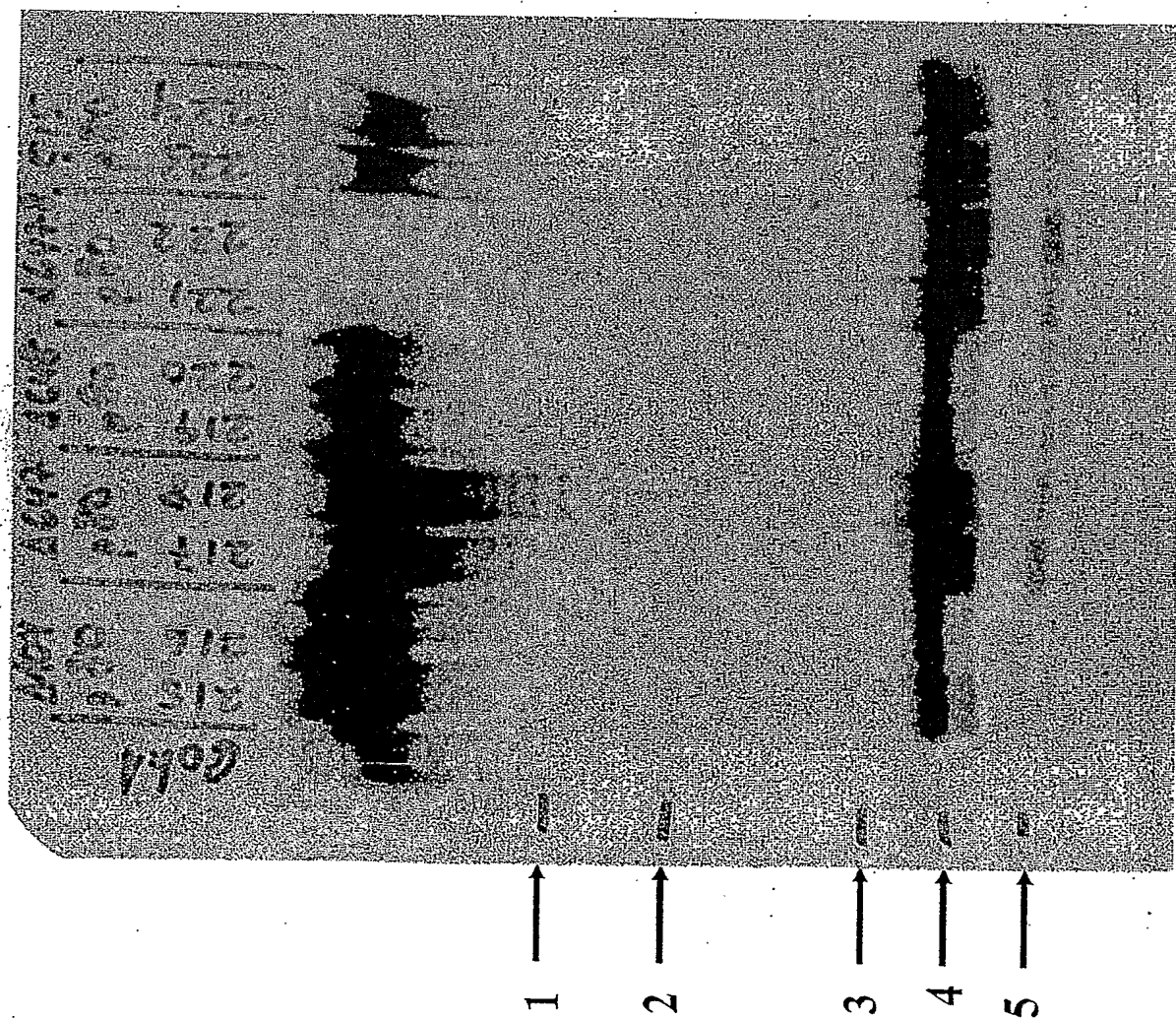


TCA precipitation of 1 ml of THB culture supernatant run on 3-8% SDS-PAGE. OD600 nm are noted above samples, "f" indicates supernatant was filtered (0.2 μ M syringe filter).

Left five samples: Coh1.

Right five samples: 179 (Δ Gbs80/pGbs80).

Figure 48: In Coh1, only the gbs80 protein and one sortase (sag0647 or sag0648) is required for polymerization



Over expression of gbs80 in various strain backgrounds (two clones each).

Total protein extract preparations.

Only the double sortase mutant does not polymerize gbs80.

Gbs80 is polymerized in the DK515 strain background (lacks adhesin island 1, adhesin island 2 is 2603-like). Presumably, sag1405&sag1406 are responsible for polymerization.

GBS STRAIN JM9030013

IEM anti-GBS80

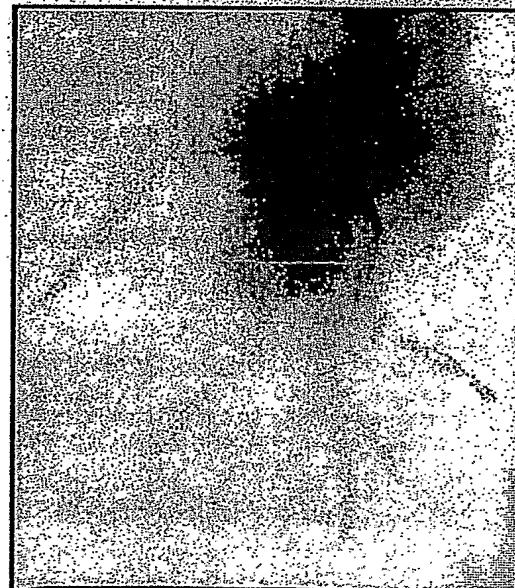
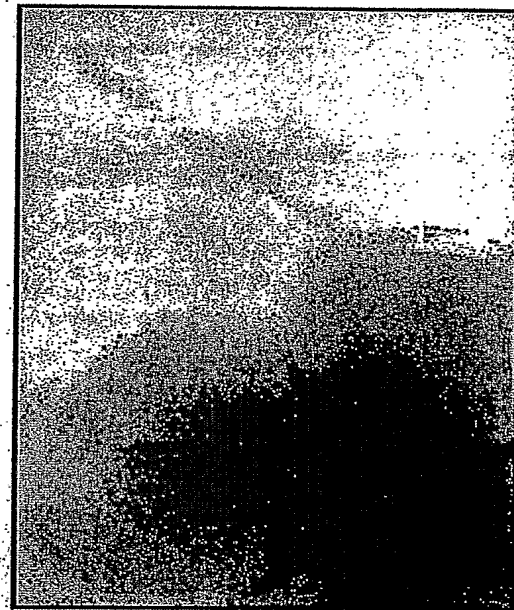
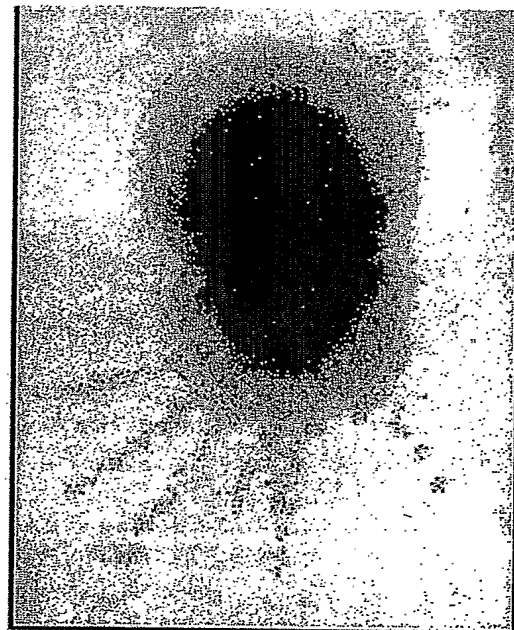
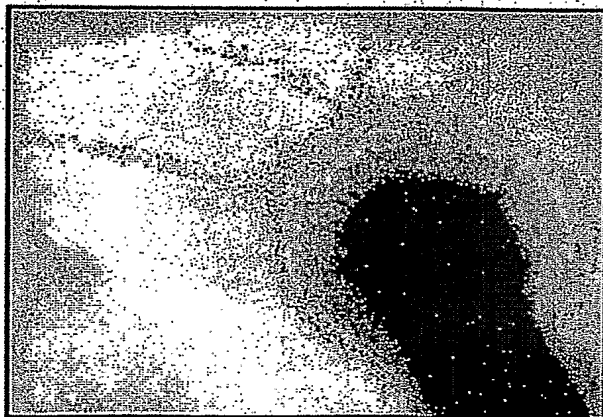
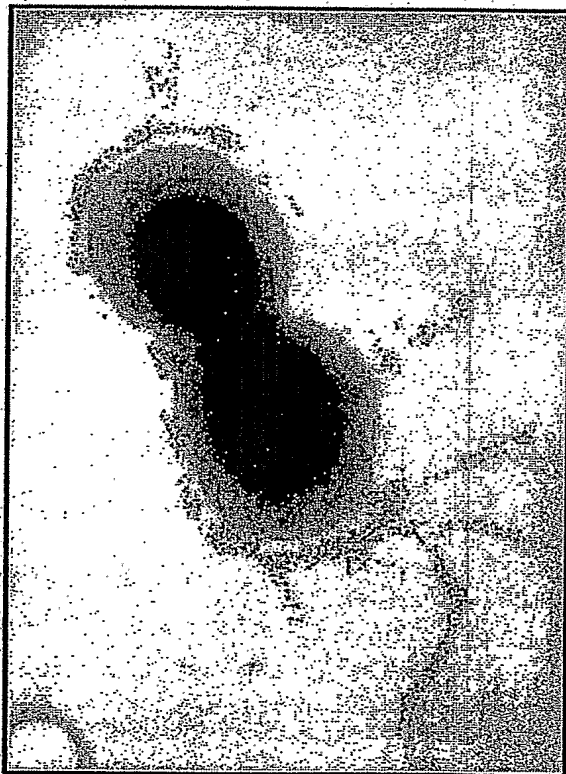
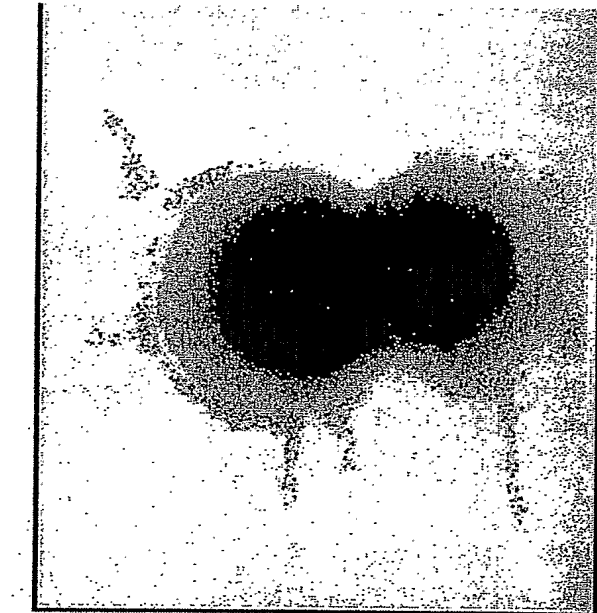
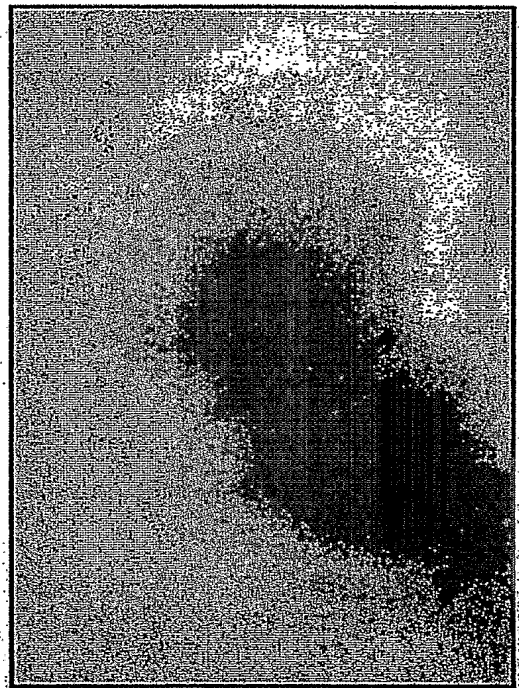
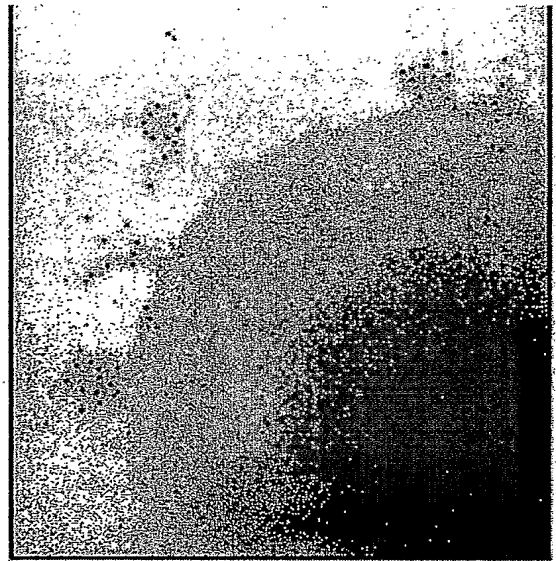
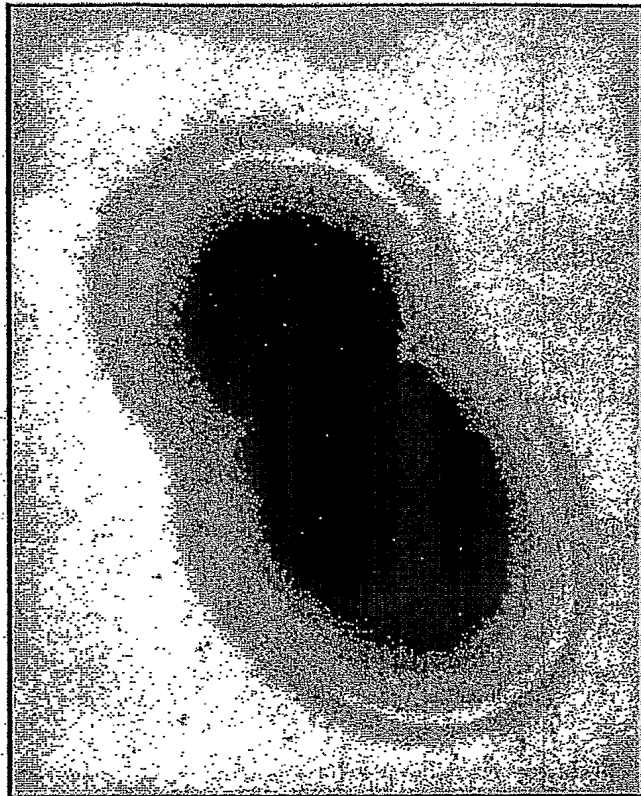
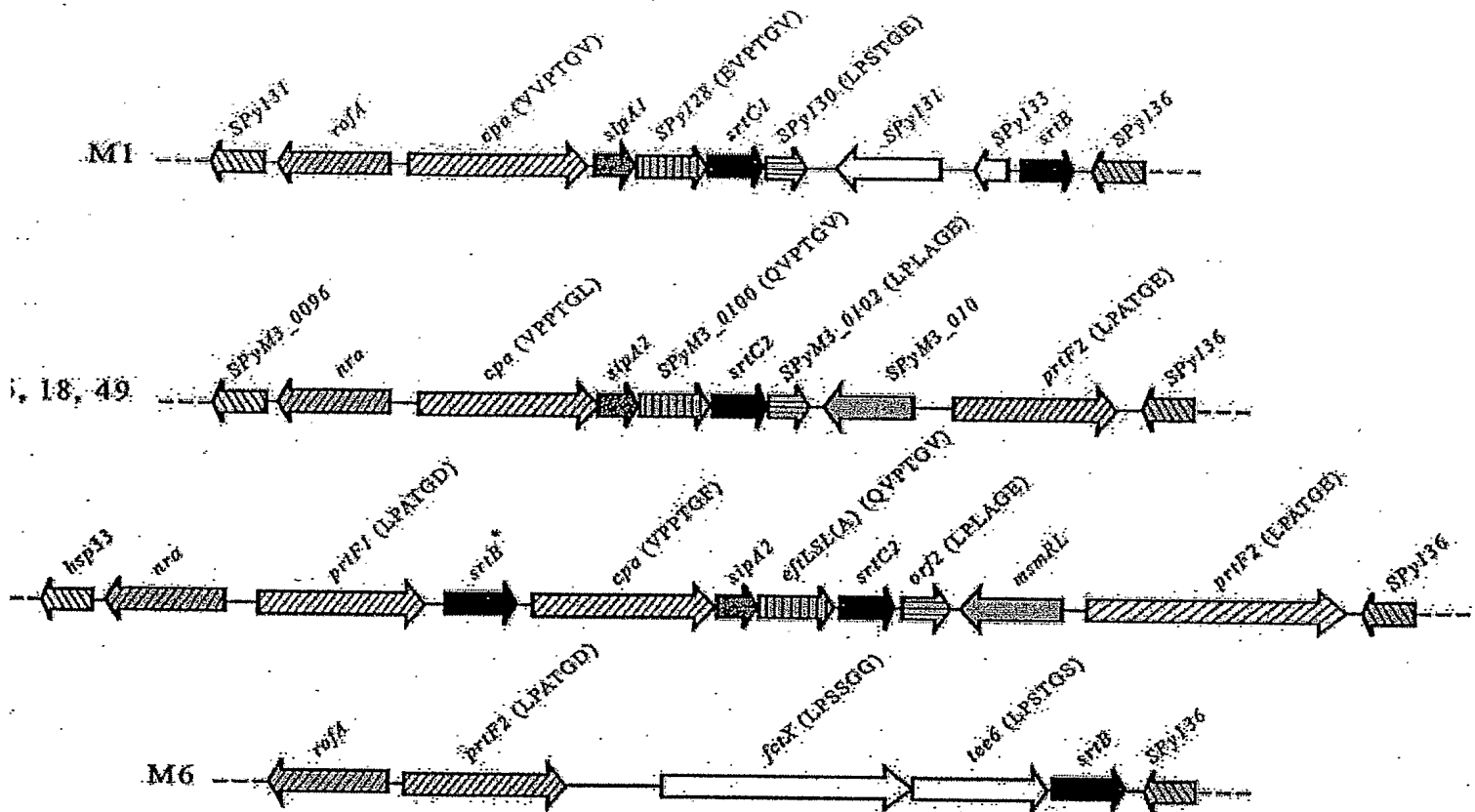


FIGURE 49



GBS STRAIN JM9030013
IEM anti-GBS104

FIGURE 50



M1	1	MAVRUKKEDKLLIVIVIC-PTTC-LTAAIT-ALIVYCC-NATNYOQYK-KKGVOFTDOLLAENS DVMKWIIVKSTHIT
M3	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCQADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M5	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCQADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M12	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCQADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M18	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCQADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M49	1	MAVTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYCQADASNFRKFKTAQOQPKFEDLLALNEDVIGRLNIPGTHL
M1	80	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M3	82	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M5	82	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M12	82	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M18	82	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M49	82	YFVOCENLEYINKAVEGSAVMSOSIFLTENHNDFTDYSLYYGHMAAGNMFGEIPKFLKKOFFNKNKATLTKR
M1	161	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE
M3	163	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE
M5	163	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE
M12	163	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE
M18	163	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE
M49	163	KLRTN-FACTTDAFTSLTENFDLVDISKN-SEINHH-KOKSV-REELTINESREVALSTCEDMTTIC-RTIVV-CAE

FIGURE 51

PCT/US05/27239

GI-19224135 1 MNKKKQKQDAPR-VSNRRP-----KQLTWTLGVFLMFLTLSSMRGAQSIFFGEEK
 ORF78 1 -----QKRDKNYGSANNKR-----ROTTIGLLKVELTFVALIG-----IVG
 GI-21909634 1 -----MOKRDKNYGSANNKR-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
 GI-28810257 1 -----MOKRDKNYGSANNKR-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
 GI-19745301 1 -----MOKRDKNYGSANNKR-----ROTTIGLLKVELTFVALIGIVGFSIRAFG
 GAS15 1 LRGEKMKTRFFPKLNTLTQRVLSKNSKRFVWTLGVFLMIFALVTSVVGAKTVFC-----

GI-19224135 53 RISEVSVEKIKSPDD--AYPWYGYDSYDSSHFYVERFVAHDLRVNNGSRSYQVYCFNL
 ORF78 39 -----
 GI-21909634 46 -AEEQSVPNKQSSVQ--DYPWYGYDSYSKGYSDYSPLNTYHNKLVNLDGSNEYQAYCFNL
 GI-28810257 46 -AEEQSVPNKQSSVQ--DYPWYGYDSYSKGYSDYSPLNTYHNKLVNLDGSNEYQAYCFNL
 GI-19745301 46 -AEEQST-----
 GAS15 58 -LVESTENAINPDSSESVRWYGYESYVRGHEVYKQFVVAHDLRVNLEGSRSYQVYCFNL

GI-19224135 111 NSHFEKKNKAFSKQFNKRVDTGCVETNYEQTPKIRGESLNNKLLSITMYNAYPKNANGYM
 ORF78 39 -----
 GI-21909634 103 TKHFFSKSDSVRSQWYKNEGTNENFIKLADKPRIEDGQLOONILRIYNGYVNDRNGIM
 GI-28810257 103 TKHFFSKSDSVRSQWYKNEGTNENFIKLADKPRIEDGQLOONILRIYNGYVNDRNGIM
 GI-19745301 52 -----
 GAS15 117 KNAFFLGSDSVKKYKRNHDCISTKEEDYMSFRTGTGDELNOKTRAVVNGHQNANGIM

GI-19224135 171 DKLEPLNAILVTOQAWWYSDSSYGN-IKTLWASELKDGKIDFEQKLMREAYSKLISDD
 ORF78 39 -----FSIRAFG
 GI-21909634 163 KCIDPLNAILVTQNAWYSDSSYISDTSKAFQCEETDLNLDSSQOLQMLRNALKNELINEK
 GI-28810257 163 KCIDPLNAILVTQNAWYSDSSYISDTSKAFQCEETDLNLDSSQOLQMLRNALKNELINEK
 GI-19745301 52 -----
 GAS15 177 EGLEPLNAIRVTOQAWWYSDNAPLISNPDSESEKRESESNLYSISQLSLNRQALNQLIDFN

GI-19224135 230 LEETSKNKLFGQSKLNTLVPODNS-----VONLLSAEYVPESPEAFQOSEPEEVQTKNT
 ORF78 46 -----AEEKSTETTKT
 GI-21909634 223 EVESLPNOVPFANYQLSIFQSSDNT-----FQNLLEAEYVDPDTPKPG-----EPPAKTEKT
 GI-28810257 223 EVESLPNOVPFANYQLSIFQSSDNT-----FQNLLEAEYVDPDTPKPG-----EPPAKTEKT
 GI-19745301 52 -----ETTKT
 GAS15 237 LATKMPKQVPDDEQLSIFQSSDNGDKYNGKYQNLLESGGVETKDETEGDPMPMPNQOTT

GI-19224135 284 SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP
 ORF78 57 SVIIRKYAEGDYSKLEGGATLRLTGEDILDFQEKVFSNGTGEKIELSNGTYTLTETSSP
 GI-21909634 275 SVIIRKYAEGDYSKLEGGATLRLAQIEGSGFQEKIFDSNKGKVELNGTYVLSEIKRP
 GI-28810257 275 SVIIRKYAEGDYSKLEGGATLRLAQIEGSGFQEKIFDSNKGKVELNGTYVLSEIKRP
 GI-19745301 57 SVIIRKYAEGDYSKLEGGATLRLAQIEGSGFQEKIFDSNKGKVELNGTYVLSEIKRP
 GAS15 297 SVIIRKYAEGDYSKLEGGATLRLTGDNVNSFQARVFSNDTSEIELSDGTYTLTETNSP

GI-19224135 344 DGYKIAEPIKFRVVKVFIQKDGSOVEMPNKEVAEPYSVAYSDMODSNYNPETFPT
 ORF78 117 DGYKIAEPIKFRVVKVFIQKDGSOVEMPNKELGSPYITLBYNDFDEEGILSTON---
 GI-21909634 335 QGYGVAPITFIVAEKVLIRNKEGQVENONKEIAEPYSVAYNDFDEEGILSTON---
 GI-28810257 335 QGYGVAPITFIVAEKVLIRNKEGQVENONKEIAEPYSVAYNDFDEEGILSTON---
 GI-19745301 117 QGYGVAPITFIVAEKVLIRNKEGQVENONKEIAEPYSVAYNDFDEEGILSTON---
 GAS15 357 AGYSIAEPIITFKVBAKVYT-IIDGKOTENPNKEIPEYSVAYNDFDEEGILSTON---

GI-19224135 404 YGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY
 ORF78 174 YAFKYGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY
 GI-21909634 393 YGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY
 GI-28810257 393 YGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY
 GI-19745301 177 YGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY
 GAS15 413 YAFKYGRFYANNKDKSSQVYCFNADLHSPFSEDGGGTIDPDLSTMKEVKYTHIAGSDLFKY

FIGURE 52

PCT/US05/27239

GI-19224135 464 ALRPRDTNPEDFLKHINKVTEKGYNKKGD--SYNGLTETQFRAATQLAIYYFTDSIDLKT
ORF78 233 ANTARDEDFQFLKHVKNVIENGYHKKGQAIPYNSLTBAQFRAATQLAIYYFTDSVDL--
GI-21909634 453 AATPRDKDADFFLKHINKVTEKGYNKKGD--TYKTLTEAQFRAATQLAIYYFTDSADLIT
GI-28810257 453 AATPRDKDADFFLKHINKVTEKGYNKKGD--TYKTLTEAQFRAATQLAIYYFTDSADLIT
GI-19745301 237 ANNPRASINDELSCQVHIVLEKGYRDDST--TYANLTSVETFAATQLAIYYFTDSVDLDN
GAS15 472 TVKPRDTDPDTFLKHINKVTEKGYREKQAIPYNSLTETQLRAATQLAIYYFTDSADLID--

GI-19224135 522 LKTYNDNKGYPHGFESWDEKTLAVTKELINYAOD--NSAPQLTNLDFFVPNNISKYQSLIGTE
ORF78 291 --TKDRLEDFHGFCDMDNDOTLGVAKKIVEYALS--DEDSKLTNLDFFVPNNISKYQSLIGTE
GI-21909634 511 LKTYNDNKGYPHGFCDKLDDEATLAVVHELTITYAED--VLEMTQNLDFFVPNNISKYQSLIGTO
GI-28810257 511 LKTYNDNKGYPHGFCDKLDDEATLAVVHELTITYAED--VLEMTQNLDFFVPNNISKYQSLIGTO
GI-19745301 295 LADY-----HGFGLTTEALNATIEYAYADRAMLPNISNLDFFVPNNISKYQSLIGTO
GAS15 531 ---KDKLNDYHGFCDMDNDSTLAVANILMEYAO--SNEPQLTDLDFVPNNISKYQSLIGTO

GI-19224135 581 YHPDDLVDVIRMEDNKKQEVIPVTHSLTVKRTVVGELGDMNGCFQFELBLNDKTSQPIVNT
ORF78 348 YHPDDLVDVIRMEDNKKQEVIPVTHSLTVKRTVVGELGDMNGCFQFELBLNDKTSQPIVNT
GI-21909634 570 YHPNELIDVISWEDNKKQPIIPITHKLTISKTVGTIADNKKREFNFEIHLKSSDQQAISGT
GI-28810257 570 YHPNELIDVISWEDNKKQPIIPITHKLTISKTVGTIADNKKREFNFEIHLKSSDQQAISGT
GI-19745301 349 YHPESLVDIIRMEDNKKQPIIPITHKLTISKTVGTIADNKKREFNFEIHLKSSDQQAISGT
GAS15 587 YHPEDLVDIIRMEDNKKQPIIPVTHSLTVKRTVVGELGDMNGCFQFELBLNDKTSQPIVNT

GI-19224135 641 LKTNNOQLVAKDGNYSFNLNKGDTIRIEGLETCGYSYTLKETE--KDYIVTVVNNVSOEAOQ
ORF78 408 LKTNNOQLVAKDGNYSFNLNKGDTIRIEGLETCGYSYTLKETE--KDYIVTVVNNVSOEAOQ
GI-21909634 630 YFTNSGELTVTDGKATFLLNDGESLIVEGLPSCGYSYEITETGASDYEVSVNGKNAPDGNA
GI-28810257 630 YFTNSGELTVTDGKATFLLNDGESLIVEGLPSCGYSYEITETGASDYEVSVNGKNAPDGNA
GI-19745301 409 YFTNSGELTVTDGKATFLLNDGESLIVEGLPSCGYSYEITETGASDYEVSVNGKNAPDGNA
GAS15 646 VKTDKTNLEFRDGNATINLHGESLITQGLETCGYSYTLKETE--SEGKVKVNSQEVANATV

GI-19224135 701 ASENVTADKEVTFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--
ORF78 468 ASENVTADKEVTFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--
GI-21909634 690 TKASVKEDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--
GI-28810257 690 TKASVKEDETVAFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--
GI-19745301 469 TKASVKEDETTFENRKDLVPPTGLTTDGAIVLWLLLLVFFGLLWVLFGRKGLKND--
GAS15 706 SKTGITTSDETFAPENNNEDVPTGVQKINGYLALETVAGISLCTHGHHTIRTKHD

FIGURE 52A

169/487

GI-19224134 WO 2006/078318 PCT/US2005/027239

GI-50913503 SYMFARGEKMNNTPLNKEAGFLVHTKRRRNFVTLVGVFTLLACAGAIGFGQVAY
MVSSYMFVARGEKMNNTPLNKEASFLAHTKRRRNFVTLVGVFTLLACAGAIGFGQVAY

GI-19224134 61 AAEKTVFNFKSPDDYPYGYDSV-----RCLEARYHNLYNLKGSREYQAYCFNITK
GI-50913503 61 AAEKTVFNFKSPDDYPYGYDAYGKEYPCYNITRYHDLRVNLNGSRSYQVYCFNITQS

GI-19224134 115 YEFPRPTYSTTNNEKKIDGSGSAFKSYAANPRVLDENLDKLEKNILNVTYNGVKSNAAGF
GI-50913503 121 NNEPSQKNSFIKNNKKIEGNCCKSFVDYAHTTKEGKE---ELEQRELSELEYNEYPNDANGY

GI-19224134 175 MNGLEDLNAITLVTONATHYYSDSAPLNDVNKMTREVRNCEISESQVTLMREALKKLIDP
GI-50913503 178 MKGGEHLNAITVTQYAVHYSDNS-QYQFETLHESEANEKISRQVTLMREALKKLIDP

GI-19224134 235 NLEATAANKIPSGYRLNIFKSENEYQNLLSAEYVPPDDPPKPGDTSEHNPKTPELDGTPI
GI-50913503 237 NLEATAVANKIPSGYRLNIFSENEAYQNLLSAEYVPPDDPPKPGETSEHNPKTPELDGTPI

GI-19224134 295 PEDPKPDESSEFALPPLMPELDGEVPEVPSESLEPALPPLMPELDGEVPEVPSESLE
GI-50913503 297 PEDPKHPDDNLEPTLPPVM-----

GI-19224134 355 PALPPLMPELDGEVPEVPSESLEPALPPLMPELDGEVPEVPSESLEPALPPLMPELDG
GI-50913503 316 -----LDGEEVPEVPSESLEPALPPLMPELDG

GI-19224134 415 EEVPEKPSVDLPFIEVPRYEFNNKDQSPLAGESGETEYITEVYGNQONPVDIDKKLPNETG
GI-50913503 343 QEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESGETEYITEVYGNQONPVDIDKKLPNETG

GI-19224134 475 FSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSE
GI-50913503 403 FSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSE

GI-19224134 535 SVEFTKDTQTGMSGOTTPOVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOVETE
GI-50913503 463 SVEFTKDTQTGMSGOTTPOIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGOTTPOIETE

GI-19224134 595 DTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATIVEDTRPKLVFHFDDNNEPKVEENREK
GI-50913503 523 DTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATIVEDTRPKLVFHFDDNNEPKVEENREK

GI-19224134 655 PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLNNKQNNKV-
GI-50913503 583 PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLNNKQNNKV

FIGURE 53

170/487

GI-19745307 1 MTQKNSY FLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQCI
ORF84 WO 2006/078318 NSYN SFLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQCAI PCT/US2005/027239
GI-28810263 1 MTQKNSYKLSFLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQCAFEIKKN
GI-21909640 1
GI-19224141 1 MTQKNSYKLSFLLSLTGFILGLLLVTIGLSGVSVGHAETRNGANKQCSFEIKKNVDQNNKP

GI-19745307 55 ----- KSQEEYNYE
ORF84 55 ----- KSQEEYNYE
GI-28810263 55 ----- KSQEEYNYE
GI-21909640 1 -----
GI-19224141 61 LPGATFSLTSKDGKGTSVQTFSTNDKGI VDAQNLQPGTYTLKEETAPDGYD KSRITVT

GI-19745307 64 VYDN ----- RNI
ORF84 64 VYDN ----- RNI
GI-28810263 64 VYDN ----- RNI
GI-21909640 1 -----
GI-19224141 121 VYENG YTKLVENPYNGEII SKAGSKDVSSSLQLENPKMSVVSKYGKTEVSSGAADFY RNI

GI-19745307 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTKNFF TAQGVSKKLYKKLSS
ORF84 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTKNFF TAQGVSKKLYKKLSS
GI-28810263 71 LDGGEHKLEIKRVDGTGKTYQG FCFQLTKNFF TAQGVSKKLYKKLSS
GI-21909640 1 ----- MSS
GI-19224141 181 AAYFKMSF EKKOK E KSETIN E GDT E V L Q L D R L N P K G I S Q D E P I T Y D S A N S P L A I G K Y H

GI-19745307 118 ----- SDEETLK
ORF84 118 ----- SDEETLK
GI-28810263 118 ----- SDEETLK
GI-21909640 4 ----- SDEETLK
GI-19224141 241 AENHQLIYTFTDYIAGLDKVQLSAELSLFLENK E V L E N T S I S N F K S T I G G Q E I T Y K G T V N

GI-19745307 125 ----- QYASKYTSNRRGDTSC
ORF84 125 ----- QYASKYTSNRRGDTSC
GI-28810263 125 ----- QYASKYTSNRRGDTSC
GI-21909640 11 ----- QYASKYTSNRRGDTSC
GI-19224141 301 VLYGNESTKESNYHTNGLSNVCGSIESYNTETGEFVWVYVNPNTNIPYATMNLWGFCR

GI-19745307 141 ----- NLKKQIAKVLTEGYPT
ORF84 141 ----- NLKKQIAKVLTEGYPT
GI-28810263 141 ----- NLKKQIAKVLTEGYPT
GI-21909640 27 ----- NLKKQIAKVLTEGYPT
GI-19224141 361 ARSNTSDLENDANTSSAELGEIOVYEVPEGEKLPSSYGVDVTKL L R T L I T A G L N G F O M

GI-19745307 157 NKS DWLN GLTENENIEVTQDAIWF
ORF84 157 NKS DWLN GLTENENIEVTQDAIWF
GI-28810263 157 NKS DWLN GLTENENIEVTQDAIWF
GI-21909640 43 NKS DWLN GLTENENIEVTQDAIWF
GI-19224141 421 TTRORI DEGN IONKAFI I KVTGKTQSGKPLVVQSNLASFRGASEYAAFTPVGGN V Y E Q

GI-19745307 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN
ORF84 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN
GI-28810263 182 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN
GI-21909640 68 TETTVPADR SYTNRNVNVSQRMKEVYQKLIDTTDID KYEDVQFDLFVPQDTN
GI-19224141 481 NEIALSPSKGSGSGKSEFTKPSITVANLRVAQLRFRKMSTDNVPLPEAAFE LRSSNGNS

GI-19745307 233 -- LQAVISVEPVIESLPITS LKPIAQNDITANN
ORF84 233 -- LQAVISVEPVIESLPITS LKPIAQNDITANN
GI-28810263 233 -- LQAVISVEPVIESLPITS LKPIAQNDITANN
GI-21909640 119 -- LQAVISVEPVIESLPITS LKPIAQNDITANN
GI-19224141 541 QKLEASSNTQGEVHFKDLTS GTYDLYETKAPKGYQQVTEKLEATVTVDTTNPAEEMVTWGS

FIGURE 54

GI-19745307 264 --- TIVVLEKPKIITYFKLYRQLPGEKEVAVDDAELK --- QINSG ---
 ORF84 WO 2006/078318 TIVVLEKPKIITYFKLYRQLPGEKEVAVDDAELK --- QINSEGO PCT/US2005/027239
 GI-28810263 264 TIVVLEKPKIITYFKLYRQLPGEKEVAVDDAELK --- QINSEGO ---
 GI-21909640 150 TIVVLEKPKIITYFKLYRQLPGEKEVAVDDAELK --- QINSEGO ---
 GI-19224141 601 PHSSVVEANKEVTILNHNKETLTFTSGKNIWENRERDORPAKIQVOLLQNGQKMPNQIQEV

GI-19745307 305 --- QEISVTWITNQLVTDERGMAYIYSVKEVDK --- NGELLEPKD --- YIKKED
 ORF84 305 --- QEISVTWITNQLVTDERGMAYIYSVKEVDK --- NGELLEPKD --- YIKKED
 GI-28810263 305 --- QEISVTWITNQLVTDERGMAYIYSVKEVDK --- NGELLEPKD --- YIKKED
 GI-21909640 191 --- QEISVTWITNQLVTDERGMAYIYSVKEVDK --- NGELLEPKD --- YIKKED
 GI-19224141 661 TKDNEVSYHEKDLPKYDANNOEYKYSVEEVNVPDGYKVSYLGNDFNTRETEFVBEQNNF

GI-19745307 349 GLTVTNTYVKPTSG --- HYDIEVTFG --- NGHIDITEDTTPDIVSGENQMK
 ORF84 349 GLTVTNTYVKPTSG --- HYDIEVTFG --- NGHIDITEDTTPDIVSGENQMK
 GI-28810263 349 GLTVTNTYVKPTSG --- HYDIEVTFG --- NGHIDITEDTTPDIVSGENQMK
 GI-21909640 235 GLTVTNTYVKPTSG --- HYDIEVTFG --- NGHIDITEDTTPDIVSGENQMK
 GI-19224141 721 NLEFGNAELNGQSGSKIIDEITLTSTFKGKKIWKNDTABNRPAIQVOLYADGVAVEGQTK

GI-19745307 394 QIEGEDS --- KPIDEVT --- ENNLIEFGKNTMPGEE
 ORF84 394 QIEGEDS --- KPIDEVT --- ENNLIEFGKNTMPGEE
 GI-28810263 394 QIEGEDS --- KPIDEVT --- ENNLIEFGKNTMPGEE
 GI-21909640 280 QIEGEDS --- KPIDEVT --- ENNLIEFGKNTMPGEE
 GI-19224141 781 FLSGSGNEWSFEFKNLKKYNGTGNDIISYKREVTVPDGYDVTYSANDIINTKREVTTOAG

GI-19745307 424 --- DGTNSNKYEEVEDSRPVDTLGSLSSSQSGDMTIEEDSATHIKFSKRD
 ORF84 424 --- DGTNSNKYEEVEDSRPVDTLGSLSSSQSGDMTIEEDSATHIKFSKRD
 GI-28810263 424 --- DGTNSNKYEEVEDSRPVDTLGSLSSSQSGDMTIEEDSATHIKFSKRD
 GI-21909640 310 --- DGTNSNKYEEVEDSRPVDTLGSLSSSQSGDMTIEEDSATHIKFSKRD
 GI-19224141 841 PKLEIBETLPLESGASGGTTTVEDSRPVDTLGSLSSSQSGDMTIEEDSATHIKFSKRD

GI-19745307 473 IDGKELAGATMELRDS SGNTISTWISDGQVDFYLMFGKTYTFVETAAPDGYEATATITFT
 ORF84 473 IDGKELAGATMELRDS SGNTISTWISDGQVDFYLMFGKTYTFVETAAPDGYEATATITFT
 GI-28810263 473 IDGKELAGATMELRDS SGNTISTWISDGQVDFYLMFGKTYTFVETAAPDGYEATATITFT
 GI-21909640 359 IDGKELAGATMELRDS SGNTISTWISDGQVDFYLMFGKTYTFVETAAPDGYEATATITFT
 GI-19224141 901 IDGKELAGATMELRDS SGNTISTWISDGQVDFYLMFGKTYTFVETAAPDGYEATATITFT

GI-19745307 533 VNEQGQVTVNGKATKGDAHIVMVDAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS
 ORF84 533 VNEQGQVTVNGKATKGDAHIVMVDAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS
 GI-28810263 533 VNEQGQVTVNGKATKGDAHIVMVDAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS
 GI-21909640 419 VNEQGQVTVNGKATKGDAHIVMVDAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS
 GI-19224141 961 VNEQGQVTVNGKATKGDAHIVMVDAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS

GI-19745307 593 SDVIIGGQC ---
 ORF84 593 SDVIIGGQC ---
 GI-28810263 593 SDVIIGGQGEVVDTTEDTQSGMTGHS ---
 GI-21909640 479 SDVIIGGQGEVVDTTEDTQSGMTGHS GSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMT
 GI-19224141 1021 SDLIIGGQGEVVDTTEDTQSGMTGHS ---

GI-19745307 602 --- QIVETTEDTQTGMHGDSCGKTEVEDTKLVQSFHFDNK
 ORF84 602 --- QIVETTEDTQTGMHGDSCGKTEVEDTKLVQSFHFDNK
 GI-28810263 619 --- GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGDSCGKTEVEDTKLVQSFHFDNK
 GI-21909640 539 GHS GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGDSCGKTEVEDTKLVQSFHFDNK
 GI-19224141 1047 --- GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGDSCGKTEVEDTKLVQSFHFDNK

GI-19745307 639 ESESNSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVISLTKKRLSSC
 ORF84 639 ESESNSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVISLTKKRLSSC
 GI-28810263 676 EPESNSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVISLTKKRLSSC
 GI-21909640 599 EPESNSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVISLTKKRLSSC
 GI-19224141 1104 EPESNSEIPKDKPKSNTSLPATGKQHNMMFFIMVTSCSLISSVFVISLTKKRLSSC

FIGURE 54A

GI-19224WO 2006/078318-MRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239

ORF80 PCT1 LBRKMRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239
 GI-21909636 1 MRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239
 GI-28810259 1 MEREMRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239
 GI-19745303 1 MRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239
 GI-13621428 1 MRKNKLLLATAILATALGTASLNQNVKAEAGVSSGOLTKRSITPCT/US2005/027239

GI-19224137 56 PNTDYTFVNFDSAPTCTESN-LPIKEGIAVN-NQIKVSYSTNDKTSQNEKQVVDVFMK
 ORF80 60 PETAFETFTLEPDMTASCHEGS-LDIRNGIVBGLDKQVTVKYKNTDKTSQNTKLAQFDFSN
 GI-21909636 56 PRADYTFKVEADSTASQKTKDGLKIPGIVNGLT-EQILSYTNTDTPDSNVKSTEFDFSK
 GI-28810259 61 PRADYTFKVEADSTASQKTKDGLKIPGIVNGLT-EQILSYTNTDTPDSNVKSTEFDFSK
 GI-19745303 56 PRADYTFKVEADSTASQKTKDGLKIPGIVNGLT-EQILSYTNTDTPDSNVKSTEFDFSK
 GI-13621428 48 PNTDYTFKVEADSTASQKTKDGLKIPGIVNGLT-EQILSYTNTDTPDSNVKSTEFDFSK

GI-19224137 114 VTFPSVGYRYRVVTEENKCTAE-CITYDDTKKTVDVYVGN-NEKGGLEPKYIVSKKCD SA
 ORF80 119 VKFPAICGVYRYMVSEKNDKED-CITYDDTKKTVDVYVGNKANNEEGFEVLIVSKKCDSS
 GI-21909636 115 VVFPGIGVYRYTVSEKQGDVE-CITYDDTKKTVDVYVGN-KEGGGFEPKPIVSKEQGTG
 GI-28810259 120 VVFPGIGVYRYTVSEKQGDVE-CITYDDTKKTVDVYVGN-KEGGGFEPKPIVSKEQGTG
 GI-19745303 116 VKFPGVGVYRYTVSEVNCNFA-CITYDDTKKTVDVYVGN-KEGGGFEPKPIVSKEQGTG
 GI-13621428 103 VTFEKEGVYRYKVTSEKIDKRVPGVSYDTTSVTVQVHVLWN-EEQQKPVATYIVGYKEGS-

GI-19224137 171 TRPPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--
 ORF80 178 TKKPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--
 GI-21909636 172 VKKPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--
 GI-28810259 177 VKKPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--
 GI-19745303 173 DKKPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--
 GI-13621428 161 VKKPIQFNNFETTSKTEKEVTGNTGDRKRAFTFTLTLOFNEYMEASSVVKIEENGQ--

GI-19224137 229 TRDVKIGEAVKFTLNDSSVLSKLPVGINVYKVEEAEANQGGYTTATLTKDC--EK
 ORF80 236 TRDVKIGEAVKFTLNDSSVLSKLPVGINVYKVEEAEANQGGYTTATLTKDC--EK
 GI-21909636 230 TRDVKIGEAVKFTLNDSSVLSKLPVGINVYKVEEAEANQGGYTTATLTKDC--EK
 GI-28810259 235 TRDVKIGEAVKFTLNDSSVLSKLPVGINVYKVEEAEANQGGYTTATLTKDC--EK
 GI-19745303 231 TRDVKIGEAVKFTLNDSSVLSKLPVGINVYKVEEAEANQGGYTTATLTKDC--EK
 GI-13621428 220 QAPVTEASIDQVTFHTLTKDGESEKVTNLFVGVDAVVTEDDYKSEKVTNINVEVSPQDGV

GI-19224137 283 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK
 ORF80 292 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK
 GI-21909636 285 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK
 GI-28810259 290 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK
 GI-19745303 285 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK
 GI-13621428 280 LSTVNLG-QEHTKTKTADIEIVVTNNRDTQVPTGVVGTLPFAVLISIVAIGGVYITRKK

GI-19224137 342 A
 ORF80 352 A
 GI-21909636 344 A
 GI-28810259 349 A
 GI-19745303 344 A
 GI-13621428 340 A

FIGURE 55

173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLFSVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLFSVVMILTMLAFNOTVLAKDSTV
ORF82	1	MLFSVVMILTMLAFNOTVLAKDSTV
GI-19745305	1	MLFSVVMILTMLAFNOTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEBITIAGSGKASFSPLTFTTVGQYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEBITIAGSGKASFSPLTFTTVGQYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEBITIAGSGKASFSPLTFTTVGQYTYRV
ORF82	61	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEBITIAGSGKASFSPLTFTTVGQYTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEBITIAGSGKASFSPLTFTTVGQYTYRV

GI-21909638	80	YOKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-28810261	86	YOKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19224139	86	YOKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
ORF82	121	YOKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19745305	92	YOKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP

GI-21909638	140	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-28810261	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-19224139	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
ORF82	181	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL
GI-19745305	152	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:====*
56	4	3:====*
58	5	2:====*
60	3	2:====*
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:=
98	0	0:

FIGURE 57

PCT/US05/27239

```

100 0 0:
102 0 0:
104 0 0:
106 1 0:=
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(54)...

```

/home/morاما/gas/pili/align/gi-50913505.pep Begin: 1 End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 6697 3452.1 9.6e-189
/home/morاما/gas/pili/align/gi-19224141.pep Begin: 48 End: 144
! gi|19224141|gb|AAL86412.1|AF447492_... 63 100 159 105.9 0.023
/home/morاما/gas/pili/align/gi-21909640.pep Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morاما/gas/pili/align/gi-13621428.pep Begin: 57 End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morاما/gas/pili/align/gi-50913506.pep Begin: 33 End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morاما/gas/pili/align/gi-13621432.pep Begin: 14 End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morاما/gas/pili/align/gi-19745301.pep Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morاما/gas/pili/align/gas15.pep Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morاما/gas/pili/align/gi-21909636.pep Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morاما/gas/pili/align/gi-28810259.pep Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morاما/gas/pili/align/gi-19224139.pep Begin: 90 End: 143
! gi|19224139|gb|AAL86410.1|AF447492_... 43 43 54 58.9 8.9
/home/morاما/gas/pili/align/gi-19745305.pep Begin: 96 End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morاما/gas/pili/align/orf82.pep Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morاما/gas/pili/align/gi-21909638.pep Begin: 84 End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List

```

gi-50913505.pep

/home/morاما/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Initl: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189
 >>/home/morاما/gas/pili/align/gi-50913505.pep (1036 aa)
 initn: 6697 initl: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189
 Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap
 (1-1036:1-1036)

10 20 30 40 50 60
 gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLPLIRRLGLQMISTQTKVIPQEIVT

FIGURE 57A

FIGURE 57B

PCT/US05/27239

```

      670      680      690      700      710      720
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLSGEEHYGITDTAELE
|||||
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLSGEEHYGITDTAELE
      670      680      690      700      710      720

      730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
|||||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
      730      740      750      760      770      780

      790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFSKDYKVDDEYTYTSLFNVKASDEAYEKYKDNEGRYS
|||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFSKDYKVDDEYTYTSLFNVKASDEAYEKYKDNEGRYS
      790      800      810      820      830      840

      850      860      870      880      890      900
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
|||||
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
      850      860      870      880      890      900

      910      920      930      940      950      960
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
|||||
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
      910      920      930      940      950      960

      970      980      990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIYQLPSSGGRGSQIFIIVGSMTA
|||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIYQLPSSGGRGSQIFIIVGSMTA
      970      980      990      1000      1010      1020

      1030
gi-50913505. TVALLFYRRQHRKKQY
|||||
gi-50913505. TVALLFYRRQHRKKQY
      1030

```

gi-50913505.pep

/home/morana/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES Initl: 63 Initn: 100 Opt: 159 z-score: 105.9 E(): 0.023

>>/home/morana/gas/pili/align/gi-19224141.pep (1161 aa)

initn: 100 initl: 63 opt: 159 Z-score: 105.9 expect(): 0.023

Smith-Waterman score: 159; 36.7% identity in 98 aa overlap

(895-990:48-144)

```

      870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKK-IV
      :| :|:| :| :| :|
gi-19224141. FILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGT
      20      30      40      50      60      70

      930      940      950      960      970      980

```

FIGURE 57C

FIGURE 57D

FIGURE 57E

initn: 149 initl: 70 opt: 86 Z-score: 71.3 expect(): 1.9
 Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
 (503-966:33-428)

```

      480      490      500      510      520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRDDAVKNSLLGVNG-----L
      |:: || :: :||::: :: :::: |
gi-50913506. NRRETVREKILITAKKLMACILAVVGLGMRVS-ALSKDDTAQLKITNIEGGPTVTL
      10      20      30      40      50      60

      530      540      550      560      570      580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      : :: | | |:: | | : : | : : | : : | : |::
gi-50913506. YKIGEGVYNTNGDSFINFK-----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
      70      80      90      100     110

      590      600      610      620      630      640
gi-50913505. LLDPNLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
      : | :: : |:: : ||| : ||| |:: : | || : :
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS-----YNGEGNLVTK
      120     130     140     150     160

      650      660      670      680      690      700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGE
      | : | :: | : | :: | |; |:: : | : |:: | : | : | : |
gi-50913506. NIDS--KSNLYGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFE
      170     180     190     200     210

      710      720      730      740      750      760
gi-50913505. EHYYGITDTABLEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQPDVLTTRKSKVNDETE
      | | :::: || | : ||::: : : : : : : | | : | : ||
gi-50913506. LPSY-----TKEAVNKTVY-----VSDNMSEGLTF--NFNSLTVEWKGMANITE
      220     230     240     250     260

      770      780      790      800      810      820
gi-50913505. ILYQKDQVQEAGKDIIDKVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTILSFNVKASDE
      | |::: | : | : | : | : : | : : | :
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      270     280     290

      830      840      850      860      870      880
gi-50913505. AYEKYKDNEGRYSEMCDSDTDYGTINQTSSEKGGGLPSNSDASVNYMADGREQKLPYKHPVI
      : | : |:: : | | : : : | | : | : | : : | : |
gi-50913506. SYKAVVNKKAIVGEEGNPNKAEFFYSNNPTKGNLYDNLDKKPKK-CNGITSKEDSK---
      300     310     320     330     340     350

      890      900      910      920      930      940
gi-50913505. QVKTVPITFTTKVDADNNQKLAGVEFELRKEDKKIVWEKGTGSGNQLNFKYLQKGKTTY
      | | | : | ||: : : | | : : : | | : | | : |
gi-50913506. IVITYQIAERKVDV-VSKTPLIGAIFGVYDTSNKLI-DIVTINKNGYAISTQVSSGK-YK
      360     370     380     390     400

      950      960      970      980      990     1000
gi-50913505. LYETKAKLGYTTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
      : | | | | : | : : : :
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
      410     420     430     440     450     460

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene

FIGURE 57F

PCT/US2005/027239/181/487

s]

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)
 initn: 65 init1: 40 opt: 78 z-score: 68.0 expect(): 2.9
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap
 (368-411:14-56)

	340	350	360	370	380	390
gi-50913505.	KGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMK--RSAPVEKFEGELE					
gi-13621432.			MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV			
			10	20	30	40

	400	410	420	430	440	450
gi-50913505.	HHKRIDYLGDNQNNPDITIDDKEDHDTSDLYRLYLDMTGKKNPLDIEVVVDKSGSMQEG					
gi-13621432.	HHELI---GDSCTCPDCHGTLTEIGSVVQRQELVFIPAQLKRINHVQHAYKCQTCSDNSL					
	50	60	70	80	90	100

gi-50913505.pep

/home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)
 initn: 52 init1: 52 opt: 73 z-score: 64.8 expect(): 4.3
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap
 (759-989:241-466)

	730	740	750	760	770	780
gi-50913505.	DSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQVQAG-KDIIDKVV					
gi-19745301.	ETIDPDFNEGKEIKYTHILGADLFSYANNPRASTNDE--LLSQVKKVLEKGYRD--DSTT					
	220	230	240	250	260	

	790	800	810	820	830	840
gi-50913505.	FTPKTTSQPKGKVTIT--FKSDYKVVDD--EYTYLSEFNVKASDEAYEKYKDNENGRYSEM					
gi-19745301.	YANLTSVEFRAATQLAIYYFTDSVDLDNLADYHGFALTEALNATKEIVAYAEADRANLP					
	270	280	290	300	310	320

	850	860	870	880	890
gi-50913505.	GDSDTDY---GTNQTSS--GKGGLPSNSDASVNYMADGREQKLPHYKHPVIQVKTIVPITFT				
gi-19745301.	NISNLDFYVPNSNKYQSLIGTQYHP--ESLVDIIRMEDKQAPIIPITHKLTISKTVTGTTI-				
	330	340	350	360	370

	900	910	920	930	940	950
gi-50913505.	KVDADNNQKKLAGVEFELRKEDKKIVWEKGTGSGN-GQLNFKYLQKQK-TYYLYETKAKL					
gi-19745301.	---AD--KKKEFNFEIHLKSSDQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI					
	390	400	410	420	430	

	960	970	980	990	1000	1010
gi-50913505.	GYTLPEN-PWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFII					
gi-19745301.	VEGLPSGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT					

FIGURE 57G

PCT/US05/27239

440 450 460 470 480 490

1020 1030

gi-50913505. VGSMTATVALLFYRRQHRKKQY

gi-19745301. GLTTDGAIIYLWLLLLVLLGLWVWLIGRKLKND

500 510 520

gi-50913505.pep

/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 43 Initn: 68 Opt: 69 z-score: 61.4 E(): 6.6

>>/home/morama/gas/pili/align/gas15.pep (762 aa)

initn: 68 init1: 43 opt: 69 Z-score: 61.4 expect(): 6.6

Smith-Waterman score: 100; 21.4% identity in 252 aa overlap
(641-873:492-739)

```

          620      630      640      650      660
gi-50913505. AKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI
          |||  :|:  :  :|:  :|  :|:  :|
gas15.pep    HIAGRDLFKYTVKPRDTPDPTFLKHIIKKVIEKGYRE-KGQAIEYSGLTETQLRAATQLAI
          470      480      490      500      510      520

```

```

        670      680      690      700      710      720
gi-50913505. DEF--KARYPNLSIYSLGVSKDINSSTASSPVVLKYLSGEEHYIGITDTAELEKTLNKI
          |  :|:  :  :|  :|:  :  :|  :|  :|  :|  :|  :|
gas15.pep    YYFTDSAEIADKDKLDYHGFDMNDSTLAVAKILVEY-AQDSNPPQLTDLDFIPNNNKY
          530      540      550      560      570

```

```

        730      740      750      760      770
gi-50913505. VEDSKLSQLGISDSLSQYVDYYDKQPDVLVT----RKSQVN---DETEILYQKDQVQEA
          :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
gas15.pep    --QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN
          580      590      600      610      620      630

```

```

        780      790      800      810      820      830
gi-50913505. GKDIIDKVVFTPKTTSQPK-GKVTILTFKSDYKVDDE-YTYTILSFNVKASDEAYEKYKDNE
          :|:  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
gas15.pep    KQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLOGLPEGYSYLVKETDSEGYKVKVNS
          640      650      660      670      680      690

```

```

        840      850      860      870      880      890
gi-50913505. GRYSMGDSDDTYGTNOT----SSGKGLPSNSDASVN-YMADGREOKLPYKHPVIOVKI
          :  :  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
gas15.pep    QEVANATVSKTGITSDETLAFENNKEPVVPTGVQDKINGYLALIVIAGISLGIWGIHTIR
          700      710      720      730      740      750

```

```

        900      910      920      930      940      950
gi-50913505. VPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTGSGNQLNFKYLQKGTYYLYET
gas15.pep    IRKHD
          760

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

FIGURE 57H

183/487

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1
 >>/home/morاما/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.8 expect(): 7.1
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:176-298)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGI VNPNDTVQTITPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||
gi-21909636. VDVYVGNKEGGGFEPKFIVSKEQGT DVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      150      160      170      180      190

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : | :|::| :||:: | : : : | : : : | : : | : : |
gi-21909636. LQKEFDFTLTINESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      200      210      220      230      240      250

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTI DNLYTWDDQ--FYLLDFISKQYEV LKTDYQSAKDSTPQTRDI
      : : : : : || :| | | :| ||: | |
gi-21909636. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESAD EIVVTNKRDTQVPTGV
      260      270      280      290      300      310

      330      340      350      360      370      380
gi-50913505. LFGEYTV EPLVMNKGHNNTINIYIRSTRLGLKPIGAAPALIQPRSF RSLTPRSTRMKRS
gi-21909636. VGT LAPFAVLSIVAIGGVIIYITKRKKA
      320      330      340

```

gi-50913505.pep
 /home/morاما/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2
 >>/home/morاما/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.7 expect(): 7.2
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:181-303)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGI VNPNDTVQTITPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||
gi-28810259. VDVYVGNKEGGGFEPKFIVSKEQGT DVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      160      170      180      190      200

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : | :|::| :||:: | : : : | : : : | : : | : : |
gi-28810259. LQKEFDFTLTINESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      210      220      230      240      250      260

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTI DNLYTWDDQ--FYLLDFISKQYEV LKTDYQSAKDSTPQTRDI
      : : : : : || :| | | :| ||: | |
gi-28810259. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESAD EIVVTNKRDTQVPTGV
      270      280      290      300      310      320

```

FIGURE 57I

PCT/US05/27239

330 340 350 360 370 380
 gi-50913505. LFGEYTVPEPLVMNKGHNNTINIIYIRSTRPIGLKPIGAAPALIQPRSFRLTPRSTRMKRS

gi-28810259. VGTLPAPFAVLSIVAIGGVIIYITKRKKA
 330 340

gi-50913505.pep
 /home/morاما/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

SCORES Initl: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9
 >>/home/morاما/gas/pili/align/gi-19224139.pep (189 aa)
 initn: 43 initl: 43 opt: 54 Z-score: 58.9 expect(): 8.9
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
 (742-796:90-143)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
					:
gi-19224139.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAY					
		:: :	: ::	:		
gi-19224139.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNIPKTPPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

gi-50913505.pep
 /home/morاما/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]

SCORES Initl: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9
 >>/home/morاما/gas/pili/align/gi-19745305.pep (195 aa)
 initn: 43 initl: 43 opt: 54 Z-score: 58.8 expect(): 9
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
 (742-796:96-149)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
					:
gi-19745305.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	70	80	90	100	110 120

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAY					
		:: :	: ::	:		
gi-19745305.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPDIPKTPPLAGEVKSLLGILSIVLLGL					
	130	140	150	160	170	180

gi-50913505.pep
 /home/morاما/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
 generated symbols 1 to: 224.

FIGURE 57J

PCT/US2005/027239
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:125-178)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
					:
orf82.pep	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	100	110	120	130	140 150

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY					
		:: :	: :: :			
orf82.pep	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGL					
	160	170	180	190	200	210

gi-50913505.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap
(742-796:84-137)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
					:
gi-21909638.	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY					
		:: :	: :: :			
gi-21909638.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:46:18 2004

! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:02.2

! Total CPU time: 0:00:02.3

! Output File: gi-50913505.fasta

FIGURE 57K

PCT/US2005/027239
 !!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
 MGAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences
 z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:=====*
50	24	5:=====*
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:==*
60	3	2:==*
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*
70	2	1:*=
72	0	0:
74	3	0:==
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=

```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(55)...

```

/home/morana/gas/pili/align/gi-50913506.pep  Begin: 1 End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454 3454 3454 1016.2 4.7e-53
/home/morana/gas/pili/align/orf84.pep  Begin: 316 End: 567
! TRANSLATE of: orf84.seq check: 7868... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-19745307.pep  Begin: 316 End: 567
! gi|19745307|ref|NP_606443.1| protei... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-21909640.pep  Begin: 202 End: 524
! gi|21909640|ref|NP_663908.1| protei... 56 81 134 75.0 1.2
/home/morana/gas/pili/align/gi-28810263.pep  Begin: 316 End: 638
! gi|28810263|dbj|BAC63201.1| protein... 56 82 134 74.7 1.3
/home/morana/gas/pili/align/orf80.pep  Begin: 49 End: 352
! TRANSLATE of: orf80.seq check: 9824... 45 69 113 70.8 2.1
/home/morana/gas/pili/align/gi-19224137.pep  Begin: 25 End: 342
! gi|19224137|gb|AAL86408.1|AF447492... 45 69 109 69.8 2.4
/home/morana/gas/pili/align/gi-19224141.pep  Begin: 277 End: 645
! gi|19224141|gb|AAL86412.1|AF447492... 73 73 118 68.9 2.7
/home/morana/gas/pili/align/gi-21909636.pep  Begin: 44 End: 344
! gi|21909636|ref|NP_663904.1| conser... 45 98 96 66.1 3.8
/home/morana/gas/pili/align/gi-28810259.pep  Begin: 49 End: 349
! gi|28810259|dbj|BAC63197.1| hypothe... 45 98 96 66.0 3.8
/home/morana/gas/pili/align/gas15.pep  Begin: 222 End: 470
! GAS15 GAS15 42 68 96 63.8 5
/home/morana/gas/pili/align/gi-13621428.pep  Begin: 17 End: 340
! gi|13621428|gb|AAK33238.1| hypothet... 41 41 87 63.6 5.2
/home/morana/gas/pili/align/gi-19224135.pep  Begin: 193 End: 462
! gi|19224135|gb|AAL86406.1|AF447492... 41 41 86 61.0 7
/home/morana/gas/pili/align/gi-50913505.pep  Begin: 503 End: 966
! gi|50913505|ref|YP_059477.1| Collag... 70 149 86 60.1 7.8
/home/morana/gas/pili/align/gi-13621430.pep  Begin: 60 End: 143
! gi|13621430|gb|AAK33240.1| hypothet... 43 67 67 59.2 8.7
/home/morana/gas/pili/align/gi-19745303.pep  Begin: 44 End: 344
! gi|19745303|ref|NP_606439.1| hypoth... 51 106 69 58.4 9.5
\\End of List

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

SCORES Initl: 3454 Inith: 3454 Opt: 3454 z-score: 1016.2 E(): 4.7e-53
 >>/home/morana/gas/pili/align/gi-50913506.pep (556 aa)
 initn: 3454 initl: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53
 Smith-Waterman score: 3454; 100.0% identity in 556 aa overlap

FIGURE 58A

PCT/US2005/027239

(1-556:1-556)

```

      10      20      30      40      50      60
gi-50913506. MTNRRETQVREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
|||||
gi-50913506. MTNRRETQVREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENV SIS
|||||
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENV SIS
      70      80      90     100     110     120

      130     140     150     160     170     180
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGLVTKNIDSKSNLYLGQTSV
|||||
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGLVTKNIDSKSNLYLGQTSV
      130     140     150     160     170     180

      190     200     210     220     230     240
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFELPSYTK EAVNKTIVYVSDNMSEGL
|||||
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFELPSYTK EAVNKTIVYVSDNMSEGL
      190     200     210     220     230     240

      250     260     270     280     290     300
gi-50913506. TFNFNSLTVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
|||||
gi-50913506. TFNFNSLTVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
      250     260     270     280     290     300

      310     320     330     340     350     360
gi-50913506. AVVNKKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKN GITSKEDSKIVYTYQIA
|||||
gi-50913506. AVVNKKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKN GITSKEDSKIVYTYQIA
      310     320     330     340     350     360

      370     380     390     400     410     420
gi-50913506. FRKVD SVSKTPLIGAIFGVYDTSNKLIDIVT FNKNGYAISTQVSSGKYKIKELKAPKGYS
|||||
gi-50913506. FRKVD SVSKTPLIGAIFGVYDTSNKLIDIVT FNKNGYAISTQVSSGKYKIKELKAPKGYS
      370     380     390     400     410     420

      430     440     450     460     470     480
gi-50913506. LNTETYEITANWVTATVKT SANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDS RPTGN
|||||
gi-50913506. LNTETYEITANWVTATVKT SANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDS RPTGN
      430     440     450     460     470     480

      490     500     510     520     530     540
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLET DIPPNTKLGELPSTG SIGTYL FKAIGSA
|||||
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLET DIPPNTKLGELPSTG SIGTYL FKAIGSA
      490     500     510     520     530     540

      550
gi-50913506. AMIGAIGIYIVKRRKA
|||||
gi-50913506. AMIGAIGIYIVKRRKA
      550
```

FIGURE 58B

189/487
PCT/US05/27239
gi-50913506.pep
/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
generated symbols 1 to: 696.
GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/orf84.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

gi-50913506.	210	220	230	240	250
	KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGKMAN----				
orf84.pep	EKEVAVDDAELKQINSEGGQEI SVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY				
	290	300	310	320	330 340
gi-50913506.	260	270	280	290	300 310
	ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNKAIVGEE				
orf84.pep	IKKEDGLTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED				
	350	360	370	380	390
gi-50913506.	320	330	340	350	
	GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIIVTYQ				
orf84.pep	SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEGGQSGDMTIE				
	400	410	420	430	440 450
gi-50913506.	360	370	380	390	400 410
	-----IAFRKVDVSKTEPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYIKE				
orf84.pep	EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMGKYTFVE				
	460	470	480	490	500 510
gi-50913506.	420	430	440	450	460 470
	LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF				
orf84.pep	TAAPDGYEIAITATFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKSGQGVIEDIEEKL				
	520	530	540	550	560 570
gi-50913506.	480	490	500	510	520 530
	YSIDSRPTGNDVKEAYIESTKALTGGTFESKSNESGTFVLETDIPNTKLGELPSTGSG				
orf84.pep	DEQCHSGSTTEIEDSKSSDVLIGGQGVIVETTEDTGTGMHGDGCKTEVEDTKLVQSFHF				
	580	590	600	610	620 630

gi-50913506.pep
/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

FIGURE 58C

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      |::: :|::: :: |: :|:::
gi-19745307. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
      290      300      310      320      330      340

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAIVGEE
      |::: :| || : |::: ::: : | |:::| |::: : || ||:
gi-19745307. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
      350      360      370      380      390

      320      330      340      350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVVYTYQ
      ::| | || :: | |::| |::: |::| |:::| |:::| |::: :| :
gi-19745307. SKPIDEV TENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQSGDMTIE
      400      410      420      430      440      450

      360      370      380      390      400      410
gi-50913506. -----IAFRKVD SVSKTFLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
      | | | |:: | || : : |::| |:: | |::: : : | | | : |
gi-19745307. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMFGKYTFVE
      460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
      || || : | |::: : ||:: |::: : : : | | | : | |
gi-19745307. TAAPDGYEIAITFTTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGSQVIDIEEKLP
      520      530      540      550      560      570

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVK EAYIESTKALTDGTTFSKSNESGSGTVLLETDPNTKLGELPSTGSIG
gi-19745307. DEQGHSGSTTEIEDSKSSDVIIGGQGQIVETTEDTQTGMHGDGCKTEVEDTKLVQSFHF
      580      590      600      610      620      630

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 56 Initn: 81 Opt: 134 z-score: 75.0 E(): 1.2
 >>/home/morana/gas/pili/align/gi-21909640.pep (656 aa)
 initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2
 Smith-Waterman score: 156; 23.9% identity in 347 aa overlap
 (232-547:202-524)

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      |::: :|::: :: |: :|:::
gi-21909640. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
      180      190      200      210      220

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAIVGEE
      |::: :| || : |::: ::: : | |:::| |::: : || ||:
gi-21909640. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
      230      240      250      260      270      280

```

FIGURE 58D

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAFFYSNNP-----TKGNTYDNL-DKKP-DKNGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::| ::| ::| ::| ::|
gi-21909640. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQSGDMTIE
          290          300          310          320          330          340

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSKTPLIGAIFGVYDTSNKLIDIVT'TNKNGYAISTQVSSGKYKIKE
          ||| || ::| || ::| ::| ::| ||| ::| ::| ::| ::| ::| ::|
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
          350          360          370          380          390          400

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
          || || : | | : : : || :: || : : : ::| | | : | ||
gi-21909640. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKGSQV-----
          410          420          430          440          450

          480          490          500          510          520          530
gi-50913506. YSIDSRTGNDVKEAYIESTKALTDTGTTFSKSNESGSGTVLLETDPNTKLGEPLSTGSIG
          || : | ::| || : | : : | : | | : | : | : | : |
gi-21909640. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
          460          470          480          490          500

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          | : : : | ::| : |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETT
          510          520          530          540          550          560

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Initl: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3

>>/home/morama/gas/pili/align/gi-28810263.pep (733 aa)

initn: 82 initl: 56 opt: 134 z-score: 74.7 expect(): 1.3

Smith-Waterman score: 155; 23.6% identity in 347 aa overlap
(232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KTTSLGSLVLSYSLTTELPSTYTKKAVNKTIVVSDNMSEGLTNNNSLTVEWKGKMAN-----
          ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
gi-28810263. EKEVAVDDAEIKQINSEGGQGEISVTWNLQVDE--KGMAYTYSVKEVDKNGELLEPKDY
          290          300          310          320          330          340

          260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLSEISPNISYKAVVNNKAIVGEE
          | :: : : | || : | : : : : | | : : | : | : | : | : |
gi-28810263. IKKEDGLTVNTFYV--KPTSGHYDIEVTFNGHIDITEDTTPDI-VSGENQMKQIEGED-
          350          360          370          380          390

          320          330          340          350
gi-50913506. GNP-----NKAFFYSNNP-----TKGNTYDNL-DKKP-DKNGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::| ::| ::| ::| ::|
gi-28810263. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQSGDMTIE
          400          410          420          430          440          450

          360          370          380          390          400          410

```

FIGURE 58E

PCT/US2005/027239

```

gi-50913506. -----IAFRKVDVSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
| | | | :: | | : : | : | | : | : : : : | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGOVKDFYLMPGKYTFVE
460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSAKSKSTTYTSDKNKATDNSEQVGWLKNGIF
| | | : | | : : : | | : | : : : : : | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQQGVTVNGKATKGDHIVMV-DAYKPTKSGSQV-----
520      530      540      550      560

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGLPSTGSIG
| | : | : : : | | : | : : : : | | : | : : | : |
gi-28810263. --IDIEKLDP-EQGHSGSTTEIEDSKSSDVLIGGQGEVVDTE--DTQSGMTGHSGST-
570      580      590      600      610      620

      540      550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
| : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGVIVETTEDTQTMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
630      640      650      660      670      680

```

gi-50913506.pep
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056
generated symbols 1 to: 352.
GETSEQ from morama, September 13, 2004 17:11.

SCORES Init1: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1
>>/home/morama/gas/pili/align/orf80.pep (352 aa)
initn: 69 init1: 45 opt: 113 Z-score: 70.8 expect(): 2.1
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap
(284-556:49-352)

```

      260      270      280      290      300      310
gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNNGFNLSFIYDSLESISPNISYKAVVN-NKAIVGEE
      :: | | : | : | : : : : : | |
orf80.pep. ATALGTASLNQNVKAETAGVVTGKSLQVTKMTYDDEEVLMPEAFTFTIEPDMTASGKE
20      30      40      50      60      70

      320      330      340      350      360      370
gi-50913506. GNPN-KAEFFYSNNPTKGNITYDNLDRKDPKNGCITSKEDSKIVTYQIAFRKVDVSVSKTP
| : : | : : : | | | : | : : : | | : : : | : :
orf80.pep. GSLDTKNGIVEGLDKQVTVKYNTRKPSQTK-LAQDFSKVKFPATGVYRYMVSEKNDK
80      90      100      110      120      130

      380      390      400      410      420
gi-50913506. LIGAIFGVYDTSNKLIDIVTTNKNG-----YAISTQ-VSSGKYKIKELKAPKGYSLN
| : | | : : : | : : | : | : : | : : : | : :
orf80.pep. KDGI---TYDDKKWTVDVYVGKANNNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
140      150      160      170      180      190

      430      440      450      460
gi-50913506. TETYEITANW-----VTATVKTSAKSKSTTYTSDKNKATDNSEQVG-----WLKNGI
| : | | : | | : | : : : : : : : : : | : |
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
200      210      220      230      240      250

470      480      490      500      510

```

FIGURE 58F

PCT/US05/27239

gi-50913506. FYSIDSRPTGNDV-----KEYAI-----ESTKALTDGTTFSKSNESGTVLLETDI
 :: | | : |:: | | : | | : | | : | | :
 orf80.pep SVMLSKLPIGINYLLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNKTKDESADIVV
 260 270 280 290 300 310

520 530 540 550
 gi-50913506. PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
 | : :: | | : | | : | : | | : | | : | | :
 orf80.pep TNKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRKKA
 320 330 340 350

gi-50913506.pep
 /home/morاما/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4
 >>/home/morاما/gas/pili/align/gi-19224137.pep (342 aa)
 initn: 69 initl: 45 opt: 109 Z-score: 69.8 expect(): 2.4
 Smith-Waterman score: 169; 26.0% identity in 334 aa overlap
 (257-556:25-342)

230 240 250 260 270 280
 gi-50913506. NKTVYVSDNMSEGLTFNFNLSLVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFI
 | : : : | : : : | | : | |
 gi-19224137. MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSIITN-FN----
 10 20 30 40

290 300 310 320 330 340
 gi-50913506. YDSLESISPNI SYKAVVN-NKAIVGEEGN-PNKAIEFFYSNNPTKGNTYDNLDDKPKDKNNG
 | : | : | : | | : | : | : | : | : | : | : | :
 gi-19224137. DDTL--LMPKTDYTFSVNPD SAATGTESNLPIKPGIAVNNQDIK-VSYNNTDKTSGKEKQ
 50 60 70 80 90 100

350 360 370 380 390 400
 gi-50913506. ITSKEDESKIVYTYQIAFRKVD SVSKTPLIGAIFGV-YDTSNKLIDIVTNNKNGYAISTQV
 : : : | : : : | : : | : | | : : | : : : : : :
 gi-19224137. VVV-DFMKVTFPSVGIYRYVVTENK---GTAEGVTYDDTKWLVDVYVGNNEKGGLPEPKY
 110 120 130 140 150 160

410 420 430 440 450
 gi-50913506. SSGKYKIKELKAPKGY--SLNTETYEITANWVTATVKT SANSKSTTYTSDKNKATDNS--
 : | : | : | : | : | : | : | : | : | : | : | : | :
 gi-19224137. IVSKKGD SATKEPIQFNNSFTTSLKIEKE-VTCNTGDKKAF TTTLTLPNEYEASSV
 170 180 190 200 210 220

460 470 480 490
 gi-50913506. ---EQVGWLKN---GIFYSI---DSR-----PTGND--VKEAYIE-----STKALTDG
 | : | : | : | : | : | : | : | : | : | : | : | : | :
 gi-19224137. VKIEENGQTRDVKIG EAYKFTLNDSQS VILSKLPVGIN YKVEEAEANQGGYTTTATLKDQ
 230 240 250 260 270 280

500 510 520 530 540 550
 gi-50913506. TTFSKSNEG---SGTVLLETDI PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVK
 : | | : : | : : : : | : | : | : | : | : | : | :
 gi-19224137. EKLSTYNL GQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITK
 290 300 310 320 330

gi-50913506. RRKA
 | : |

FIGURE 58G

PCT/US05/27239

194/487

gi-19224137. RKKA
340

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7
 >>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
 initn: 73 init1: 73 opt: 118 z-score: 68.9 expect(): 2.7
 Smith-Waterman score: 174; 23.9% identity in 406 aa overlap
 (115-483:277-645)

```

          90      100      110      120      130      140
gi-50913506. VSLTETGPTSQEITTIANGINTGKIKPFSTENVISISNGTATYNARGASVY--IALLTGAT
          ||:||||| :| ::: : : :| |
gi-19224141. IYTFTDYIAGLDKVLQSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTNVNVLGYNE
          250      260      270      280      290      300

          150      160      170      180      190
gi-50913506. DGRTYNPILLAASYNAGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVGTG-----T
          : : | | : | | |:: : | :: | :: :::| | : : | |
gi-19224141. STKESNYITNGLSNVG-GSIESYNTETGEFVWYVYVNPRTNIPYATMNLWGFGGRARSNT
          310      320      330      340      350      360

          200      210      220      230      240      250
gi-50913506. ID---DVNKKTTSLGSVLSYSITF--ELPSYTKAVNKTFTVSDNMSEGLTFNFNLSLTVE
          | | :| ::| | : | : :||| | :| : :| : : || : :|
gi-19224141. SDLENDANTSSAELGEIQVYEVPEGEKLPSSYGVDTVTKLTLRTD-ITAGLNGGFQ-----
          370      380      390      400      410

          260      270      280      290      300      310
gi-50913506. WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVG
          |:: :| :: || | :|| : : : | | :| :: ::
gi-19224141. -----MTKRQRIDFG--NNIQNKAFIIKV--TGKTDQSGKPLVVQSNLAS
          420      430      440      450      460

          320      330      340      350      360
gi-50913506. EEGNPNKAEFFYSNNPTKGNTRY--DNLDDKPKDKNGITSKEDSKIVYTY-----QIAF
          :| : | | :| :||| ::: :|:|:| :| :| | | :| :|
gi-19224141. FRGASEYAAF-----TPVGGNVYFQNEIALSPSKGSGSGKSEETKPSITVANLKRVAQLRF
          470      480      490      500      510

          370      380      390      400      410      420
gi-50913506. RKVDSVSKTPLIGAFGVYDTSNKLIDI-VTINKNGYATSTQVSSGKYKIKELKAPKGYG
          :| :| ::::| | | : :::: : :::| :| : :::| | :| | | | |
gi-19224141. KKM-STDNVLPLEAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQ
          520      530      540      550      560      570

          430      440      450      460
gi-50913506. -----LNTETYEIT-----ANWVT--ATVKTSANSKSTTYTSDKNKATDNSEQVGWLN
          | | | : | :| : :|:| | | :| :| :| :| :| :| :| :|
gi-19224141. QVTEKLATVTVDTKPAEMVTWGSPPHSSVKVEAN-KEVTIVNHKETLTFSGKKI-WEND
          580      590      600      610      620      630

          470      480      490      500      510      520
gi-50913506. GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDPNTKIGELPSTG
          | :| : :| :
gi-19224141. ----RPDQRFPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQYKYSVEE
          640      650      660      670      680

```

FIGURE 58H

gi-50913506.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8
>>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:44-344)

```

      270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
      :: : ::|::: : |::| ::: :
gi-21909636. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
      20      30      40      50      60      70

      330      340      350      360
gi-50913506. KGNTYDNLDDKPKDGNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
      :|:| |:|: || |:| :| |||: |::| || :||:
gi-21909636. SGKTKDGLEIKPGIVNGLTEQII SYTNTDKPDSKVKST-EFDFSKVVFPGIGVRYTVSE
      80      90      100      110      120

      370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SINTETY
      | : | : ||:: :|: : ||:| :: :|: : ::| |::| :|:|
gi-21909636. KQ--GDVEGITYDTPKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
      130      140      150      160      170      180

      430      440      450      460
gi-50913506. EITANWVTATVKTSA-----SKSTTYTSDK-----NKATDNSE-QVGW-----LKNGI
      :: | | : : : ::||:: :|: :|::: :|:| |||
gi-21909636. KVKKNVSGNTGELQKEFDFTLTNNESTNFKKDQIVSLQKNEKFEVKIGTPYKFKLKNGE
      190      200      210      220      230      240

      470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----
      ::|: |:| | :|::| || :|:|::| : :|:
gi-21909636. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGPGQSKMYQLDMEQKTDESADIEV
      250      260      270      280      290      300

      520      530      540      550
gi-50913506. IPNTIKLGELPSTGSIGTYLEKAIGSAAMIGAIGIYIVKRRKA
      : | : ::| || :|| :|: | : ||:: |||:|:|
gi-21909636. VTNKRDTQVP-TGVVGT LAPFAVL SIVAIGGV-IYITKRRKA
      310      320      330      340

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8
>>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:49-349)

FIGURE 58I

```

270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: : |:: : :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
20      30      40      50      60      70

330      340      350      360
gi-50913506. KGNTYDNLDDKKPKDGNIGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
:|:| |:|: || |:| :| |||: |::| || :||:
gi-28810259. SGKTKDGLLEIKPGIVNGLTEQIISYTNIDKPD SKVKST-EFDFSKVVFPGIGVYRYTVSE
80      90      100      110      120      130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVITNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
| : | : |||:: :|: : ||:| : : : | : : :| | : : | : | :
gi-28810259. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTVDVKKPVNFNNSFATTSL
140      150      160      170      180      190

430      440      450      460
gi-50913506. BITANVWTATVKTSA-----SKSTTYTSDK-----NKATDNSE-QVGW-----LKNIGI
: : | | : : : :|::: :|: :|:::| :| :| |||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTINESTNFKKDQIVSLQKGNKFEVKIGTPYKFKLKNGE
200      210      220      230      240      250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
:|:| |:| | :|:| || :| :|::| : :|
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTASLKEGDGQSKMYOLDMEQKTDSEADEIV
260      270      280      290      300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFAIGSAAMIGAIGIYIVKRRKA
: | : :| || :|| | : | : ||:: ||:|:|
gi-28810259. VTNKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRRKA
310      320      330      340

```

gi-50913506.pep
/home/morana/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Initl: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5
>>/home/morana/gas/pili/align/gas15.pep (762 aa)
initn: 68 initl: 42 opt: 96 z-score: 63.8 expect(): 5
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap
(283-535:222-470)

```

260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLES-ISPNISYKA--VVNN--K
                ||:: :|::: |:|::: | :
gas15.pep     VWYYS DNAPISNPDESFKRESESNLVSTS QLSLMRQALKQLIDPNLATKMPKQVPDDFQL
200      210      220      230      240      250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTYDNLDDKKPKDGNIGIT'SKEDSKIVYTYQIAFRK--V
:| | : :| : |:| :|: | | : : : | : :|
gas15.pep     SIFESDKGDYKNGYQNLLSGGLVPT---KPPTPGDPMPPNQPPQ---TTSVLIRKYAI
260      270      280      290      300

370      380      390      400      410      420

```

FIGURE 58J

FIGURE 58K

PCT/US05/27239
 gi-13621428. E-KTTKGGQAFVQT---EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTTEDDYKSEKYT
 220 230 240 250 260

 500 510 520 530 540
 gi-50913506. T-----DGT-----FSKSNESGTVLLETDPNTKLGELPSTGSIGTYLFKAIGSAA
 | ||:: | :| : : | | |:|: :: : : |:| :|
 gi-13621428. TNVEVSPQDGAVKNIAGNSTEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVA
 270 280 290 300 310 320

 550
 gi-50913506. MIGAIGIYIVKRRKA
 : ||: |:|:::|
 gi-13621428. VGGAL--YFVKKNA
 330 340

gi-50913506.pep
 /home/morاما/gas/pili/align/gi-19224135.pep

gi|19224135|gb|AAL86406.1|AF447492_3 Cpa [Streptococcus pyogenes]

SCORES Init1: 41 Initn: 41 Opt: 86 z-score: 61.0 E(): 7
 >>/home/morاما/gas/pili/align/gi-19224135.pep (756 aa)
 initn: 41 init1: 41 opt: 86 Z-score: 61.0 expect(): 7
 Smith-Waterman score: 101; 19.6% identity in 306 aa overlap
 (243-535:193-462)

 220 230 240 250 260 270
 gi-50913506. SLTFELPSYTKAEVNTKTVYVSDNMSEGLTFNFSNLTVEWKGKMANITEDGSVMVENTKIG
 ::: : | :: |:|: |:|:
 gi-19224135. PKNANGYMDKIEPLNAILVTQQAVWYYSDDSYGNIKTLWASEL----KDGKIDFEQVKL-
 170 180 190 200 210

 280 290 300 310 320 330
 gi-50913506. IAKEVNNNGFNLSPFIYDSLESISPNI SYKAVVNNKAIVGEEGNPN--KAIEFFYSNNPTKGN
 :|: : ::| |:| | | :: | : :: | |:|: : |: |:
 gi-19224135. -MREAYS----KLISDDLEETSKNKL PQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQ
 220 230 240 250 260 270

 340 350 360 370 380
 gi-50913506. TYDNLDDKPKDNGGITSKEDSKIVYTY-QIAFRKVDSVSKTPLIGAIFGVYDTSNKLIDI
 : | : |:| |:| | : : |: : | | |:| |:| |:
 gi-19224135. S---PEPP---VQTKTSVIIRKYAEGDYSKLLEGATLRLTGE--DILDFQEK--V
 280 290 300 310

 390 400 410 420 430 440
 gi-50913506. VTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTET-YETANWVTATVKTSA NSKSTT-
 :| | | ::|:| |:| | :| |:| |:| |:| |:| |:| |:
 gi-19224135. FQSNGTGEKI--ELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDG SQVENPNK
 320 330 340 350 360 370

 450 460 470 480 490
 gi-50913506. -----YTSKDNKATDNSEQVG---WLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTF
 |: : : ::|: : : | | :: :::| : : | :
 gi-19224135. EVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKKDKSSQVYCFN-----ADLHSP
 380 390 400 410 420 430

 500 510 520 530 540 550
 gi-50913506. SKSNEGSGTVLLETDPNTKLGELPSTGSIGTYLFKAIGSAA MIGAIGIYIVKRRKA
 :|:|:|:|: : | | | | : | : |:|
 gi-19224135. PESEDGGGTI--DPDISTMK--EVKYTHTAGSDLFKYALRPDTPN PEDFLKHKKVIEKG
 440 450 460 470 480

FIGURE 58L

199/487

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDCLKTLKTYNNGKGYHGFESMDEKTLAVTK
 490 500 510 520 530 540

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M
 GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8
 >>/home/morana/gas/pili/align/gi-50913505.pep (1036 aa)
 initn: 149 init1: 70 opt: 86 z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
 (33-428:503-966)

```

      10      20      30      40      50      60
gi-50913506. NRRETVREKILITAKKMLACLAAILAVVGLGMRVS-ALSKDDTAQLKITNIEGGPTVTL
      |:: || :: :||::: :: ::::|
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRRDDAVKNSLLGVNG-----L
      480      490      500      510      520

      70      80      90      100     110
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQBIT-TIANGINTGKIKPFSTEN
      : :: | | :|: | | : :| : : :| :|::
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      530      540      550      560      570      580

      120     130     140      150      160
gi-50913506. VVISNGTATYNARGASVYIALITGAT-----DGRTYNPILLAAS-----YNGEGNLTVK
      : | : : :|: : ||| : ||| |:: : | || : :
gi-50913505. LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSG
      590      600      610      620      630      640

      170     180     190     200     210
gi-50913506. NIDS--KSNLYLGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
      | :| ::| :| :|| :| :|:: :| :|:| :| :| :| :|
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLSCGE
      650      660      670      680      690      700

      220     230     240     250     260
gi-50913506. LPSY-----TKEAVNKIVY-----VSDNMSEGLTF--NFNSLTVEWKGKMANITE
      | :|:| | :|:|:| :|:|:| :| :| :| :| :| :| :|
gi-50913505. EHYYGITDTAELEKTNKLVESKLSQLGISDLSQYVDYDQKPDVLVTRKSKVNDETE
      710      720      730      740      750      760

      270     280     290
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      | :|:| :| :|:| :| :| :| :| :| :| :|
gi-50913505. ILYQKDQVQEAGKDIIDKVFTPKTTSQPKGVTLTFKSDYKVDDEYTYTLSTFNVKASDE
      770      780      790      800      810      820

      300     310     320     330     340     350
gi-50913506. SYKAVVNNKAIVGEEGNENKAEFFYSNNPTKGNTYDNLDDKPKDK--GNGITSKEDSK----
      :| :|:| :| :| :| :| :| :| :| :| :| :| :| :|
gi-50913505. AYEKYKDNEGRYSEMGSDTDYGTNTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI
      830      840      850      860      870      880

      360     370     380     390     400
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKL--DIVTTNKNGYAISTQVSSGK-YK
      | | :| ||| :| :| :| :| :| :| :| :| :| :| :|

```

FIGURE 58M

PCT/US05/27239

gi-50913505. QVKTVPITFTKVDADNNQKLAGVEFELRKEDKKIVWEKGTGSGNQLNFKYLQKGKTTY
890 900 910 920 930 940

410 420 430 440 450 460
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
: | | | : : : :
gi-50913505. LYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
950 960 970 980 990 1000

gi-50913506.pep

/home/morاما/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 67 Opt: 67 z-score: 59.2 E(): 8.7
>>/home/morاما/gas/pili/align/gi-13621430.pep (215 aa)
initn: 67 init1: 43 opt: 67 Z-score: 59.2 expect(): 8.7
Smith-Waterman score: 67; 27.8% identity in 90 aa overlap
(433-522:60-143)

410 420 430 440 450 460
gi-50913506. VSSGKYKIKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQV
: | | | : | | | : : : :
gi-13621430. TASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLNSVT-TSVKGNKTSFEQL
30 40 50 60 70 80

470 480 490 500 510 520
gi-50913506. GWLKNIGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKILGE
: : | : : | : : | : : : : | | : | : : : :
gi-13621430. TFSEVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLY---NEQSGA---LETNLVSNKILGE
90 100 110 120 130 140

530 540 550
gi-50913506. LPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

gi-13621430. TEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKRNIGILPSTGEMVSYSALGIVLVATIT
150 160 170 180 190 200

gi-50913506.pep

/home/morاما/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]

SCORES Init1: 51 Initn: 106 Opt: 69 z-score: 58.4 E(): 9.5
>>/home/morاما/gas/pili/align/gi-19745303.pep (344 aa)
initn: 106 init1: 51 opt: 69 Z-score: 58.4 expect(): 9.5
Smith-Waterman score: 129; 24.0% identity in 308 aa overlap
(298-556:44-344)

270 280 290 300 310 320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
: : : : : : : : : : : :
gi-19745303. ATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDKVLMPKADYTFKVE---ADDNA
20 30 40 50 60 70

330 340 350 360 370
gi-50913506. KGNTYDNLDPKPKNGI-----TSKEDSKIVYTYQIAFRKVDVSVSKT--P
| : | | : | | : | | : | | : | | : | |

FIGURE 58N

```
! Distributed over 1 thread.
!   Start time: Wed Sep 15 18:45:54 2004
! Completion time: Wed Sep 15 18:46:02 2004
! CPU time used:
!   Database scan: 0:00:00.1
! Post-scan processing: 0:00:01.9
!   Total CPU time: 0:00:02.0
! Output File: gi-50913506.fasta
```

FIGURE 580

!!SEQUENCE LIST 1.0

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences:

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	1	5:= *
46	7	5:====*==
48	21	5:====*=====
50	7	5:====*==
52	2	4:== *
54	4	3:==*==
56	3	3:==*
58	0	2: *
60	4	2:==*==
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:===
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59

102 PCT/US2005/027239
 104 0 0:
 106 0 0:
 108 0 0:
 110 0 0:
 112 0 0:
 114 0 0:
 116 0 0:
 118 0 0:
 >120 1 0:=-

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(55)...
/home/morama/gas/pili/align/gi-13621430.pep	Begin: 1	End: 215			
! gi 13621430 gb AAK33240.1 hypothet... 1338	1338	1338	233.9	1.8e-09	
/home/morama/gas/pili/align/gi-19745305.pep	Begin: 1	End: 193			
! gi 19745305 ref NP_606441.1 hypoth... 163	243	273	82.2	0.5	
/home/morama/gas/pili/align/gi-28810261.pep	Begin: 2	End: 187			
! gi 28810261 dbj BAC63199.1 hypothe... 164	239	268	81.5	0.55	
/home/morama/gas/pili/align/gi-19224139.pep	Begin: 2	End: 187			
! gi 19224139 gb AAL86410.1 AF447492.... 164	236	265	81.0	0.57	
/home/morama/gas/pili/align/orf82.pep	Begin: 30	End: 222			
! TRANSLATE of: orf82.seq check: 4296... 163	235	264	81.0	0.58	
/home/morama/gas/pili/align/gi-21909638.pep	Begin: 2	End: 181			
! gi 21909638 ref NP_663906.1 hypoth... 164	239	261	80.5	0.62	
/home/morama/gas/pili/align/gi-19745303.pep	Begin: 84	End: 183			
! gi 19745303 ref NP_606439.1 hypoth... 121	121	126	61.4	6.7	
/home/morama/gas/pili/align/gi-13621428.pep	Begin: 6	End: 174			
! gi 13621428 gb AAK33238.1 hypothet... 58	86	122	60.9	7.2	
/home/morama/gas/pili/align/gi-19224137.pep	Begin: 93	End: 201			
! gi 19224137 gb AAL86408.1 AF447492.... 88	88	119	60.4	7.5	
/home/morama/gas/pili/align/gi-50913503.pep	Begin: 549	End: 625			
! gi 50913503 ref YP_059475.1 Fibron... 73	73	117	60.4	7.6	
/home/morama/gas/pili/align/gi-19224134.pep	Begin: 631	End: 697			
! gi 19224134 gb AAL86405.1 AF447492.... 73	73	115	60.1	7.8	

\\End of List

gi-13621430.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09
 >>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)
 initn: 1338 init1: 1338 opt: 1338 Z-score: 233.9 expect(): 1.8e-09
 Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap
 (1-215:1-215).

gi-13621430.	10	20	30	40	50	60
	MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI					
gi-13621430.	MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI					
	10	20	30	40	50	60
gi-13621430.	70	80	90	100	110	120
	EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV					
gi-13621430.	EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV					

FIGURE 59A

FIGURE 59B

(8-213-2-187) US 05/27239

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSEVAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-28810261.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : :: |::|::| |::| |::| |::| :::|::| : | |::|:
gi-28810261. ESIDAMKTIEE---ITTAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKRNGI
      :|| |::| :| :::| : |::| : || : : | |::| : |::| :
gi-28810261. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::| : :| : |||:
gi-28810261. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLSKRL
      160     170     180

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```

SCORES  Init1: 164  Initn: 236  Opt: 265  z-score: 81.0  E(): 0.57
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303; 30.6% identity in 206 aa overlap
(8-213-2-187)

```

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSEVAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-19224139.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : :: |::|::| |::| |::| |::| :::|::| : | |::|:
gi-19224139. ESIDAMKTIEE---ITTAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKRNGI
      :|| |::| :| :::| : |::| : || : : | |::| : |::| :
gi-19224139. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::| : :| : |||:
gi-19224139. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLSKSL
      160     170     180

```

FIGURE 59C

PCT/US05/27239

gi-13621430.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58
Smith-Waterman score: 304; 30.5% identity in 213 aa overlap
(1-213:30-222)

```

gi-13621430.      10      20      30
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTA
orf82.pep      LLFQRVKIFLLTIVLSLSVLFKNNERRRLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
                  10      20      30      40      50      60

gi-13621430.      40      50      60      70      80      90
                  SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLNSVTTSVKNGKTSFEQLTFFS
orf82.pep      QTSISVENVLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT
                  70      80      90      100     110

gi-13621430.      100     110     120     130     140     150
                  EVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF
orf82.pep      TVGQYTYRVYQKPSQNKDYQADTTVFVDLVVYTYDED-GTLVAKVISRRAGDEEKSAITF
                  120     130     140     150     160     170

gi-13621430.      160     170     180     190     200     210
                  KQEYSEKTPEPHQPDTEKEKPKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKL
orf82.pep      KPKRLVKPIPPROPNI-----PKTP-----LPLAGEVKSLLGILSIVLLGLLVLLYV-KKL
                  180     190     200     210     220

gi-13621430.      KTSK
orf82.pep      KSRL

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31
5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62
Smith-Waterman score: 302; 31.5% identity in 200 aa overlap
(14-213:2-181)

```

gi-13621430.      10      20      30      40      50      60
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-21909638.      :| : : :| : : :| : : :| : : :| : : :| : : :
                  MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL

```

FIGURE 59D

PCT/US05/27239

```

                70      80      90      100      110      120
gi-13621430. EALDKESPLPNSVTTSVKGNKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEV
|::| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
                50      60      70      80      90

                130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQYSEKTPPEHPQDPTTEKEKPQKKRNGI
:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-21909638. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPROPNI-----PKTP----
                100      110      120      130      140

                190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-21909638. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
                150      160      170      180

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

```

SCORES   Init1: 121   Initn: 121   Opt: 126   z-score: 61.4   E(): 6.7
>>/home/morana/gas/pili/align/gi-19745303.pep           (344 aa)
  initn: 121 init1: 121 opt: 126 Z-score: 61.4 expect(): 6.7
Smith-Waterman score: 126;    27.0% identity in 100 aa overlap
(59-155:84-183)

```

```

                30      40      50      60      70      80
gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTSVKNGN-KTSF
|::| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745303. LMPKADYTFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFD
                60      70      80      90      100      110

                90      100      110      120      130      140
gi-13621430. EQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETE
:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19745303. ANVKFPGVGVIYRTVSEVNGNKAGIAYDSQQWTVDVYVNRDGGFEAKYIVSTEGGQSD
                120      130      140      150      160      170

                150      160      170      180      190      200
gi-13621430. KSELIFKQYSEKTPPEHPQDPTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLY
|: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-19745303. KKPVLFRNFFDITSLKVTKKVTGNTGEHQRSFSFTLLTPNECFEKGQVNNILQGGETKK
                180      190      200      210      220      230

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

```

SCORES   Init1: 58   Initn: 86   Opt: 122   z-score: 60.9   E(): 7.2
>>/home/morana/gas/pili/align/gi-13621428.pep           (340 aa)
  initn: 86 init1: 58 opt: 122 Z-score: 60.9 expect(): 7.2
Smith-Waterman score: 135;    29.1% identity in 172 aa overlap

```

FIGURE 59E

(8-159) PCT/US2005/027239

```

      10      20      30      40      50
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-
      :| | | | :|::|::| : : |::| | : | | :
gi-13621428. MKLRHLLLTGAALTSFAAT-TVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIE
      10      20      30      40      50

      60      70      80      90      100
gi-13621430. ---VIEALDK-----ESPLPN-SVTTSVKNGKTSFEQLTFSEV-----GQYHYKI-H
      | | :| :|::| : | |::| : : | | | | | | | | : | | :
gi-13621428. PDDTVNEDGNKFKGVALNTPMTKVITYNSDKGGSNTKTAEFDFSEVTFEKPQGVYVYKVT
      60      70      80      90      100      110

      110      120      130      140      150      160
gi-13621430. QLLGKNSQYHYDETVYEVVYVLYNE-QSCALETNLVSNKLGETEKSELIFKQEYSEKTP
      : : | | | | :|::| | : : | : | | | | : | | :
gi-13621428. EKIDKVPQVSYDTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQFKNSLDSTTL
      120      130      140      150      160      170

      170      180      190      200      210
gi-13621430. EPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
gi-13621428. TVKKKVSQGGDRSKDFNGLTLKANQYKASEKVMIEKTTKGGQAPVQTEASIDQLYHF
      180      190      200      210      220      230

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

```

SCORES   Initl: 88   Initn: 88   Opt: 119   z-score: 60.4   E(): 7.5
>>/home/morama/gas/pili/align/gi-19224137.pep   (342 aa)
initn: 88 initl: 88 opt: 119 Z-score: 60.4 expect(): 7.5
Smith-Waterman score: 119;   29.7% identity in 111 aa overlap
(72-176:93-201)

```

```

      50      60      70      80      90
gi-13621430. DVATNKQSSDIDETFMFVIEALDKESPLPNVTTSVKNGK-----TSFEQLTFSEVGQY
      |::|::| |::| |::| | |
gi-19224137. SVNPDSAATGTESNLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDPMKVTFFPSVGII
      70      80      90      100      110      120

      100      110      120      130      140      150
gi-13621430. HYKIHOLLGKNSQYHYDETVYEVVYVLYNEQSCALETNLVSNKLGETEKSELI-FKQEY
      : : : | | : | : | | : | : | | : | : | | :
gi-19224137. RYVVTENKGTAEQVYDDTKWLVDVYVGNNEKGG-LEPKYIVSKKGD SATKEPIQFNNSF
      130      140      150      160      170      180

      160      170      180      190      200      210
gi-13621430. SEKTPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | | : : | : : |
gi-19224137. -ETTSLKIEKEVTGNTGDHKAFTFTLTLPNEYEASSVVKIEENGQTKDVKIGEAYKF
      190      200      210      220      230      240

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-50913503.pep

gi|50913503|ref|YP_059475.1| Fibronectin-binding protein [Streptococcus pyogenes MGAS10394]

FIGURE 59F

PCT/US05/27239

SCORES Init1: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6
 >>/home/morana/gas/pili/align/gi-50913503.pep (627 aa)
 initn: 73 init1: 73 opt: 117 Z-score: 60.4 expect(): 7.6
 Smith-Waterman score: 118; 28.7% identity in 87 aa overlap
 (129-215:549-625)

```

      100      110      120      130      140      150
gi-13621430. KIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEK
                        || || | | : : | : : : |
gi-50913503. IETEDTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATVV---EDTRPKLVFHFDDNNEP
      520      530      540      550      560      570

      160      170      180      190      200      210
gi-13621430. TPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | : : : || | : : || : || : : : : || : : : : : : | : : : |
gi-50913503. KVEEN---REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQSNKKV
      580      590      600      610      620

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492_2 protein F [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8
 >>/home/morana/gas/pili/align/gi-19224134.pep (698 aa)
 initn: 73 init1: 73 opt: 115 Z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 115; 27.4% identity in 73 aa overlap
 (143-215:631-697)

```

      120      130      140      150      160      170
gi-13621430. DETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEK
                        | : : | : : : | | : : ||
gi-19224134. VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDDNNEPKVEEN-----REK
      610      620      630      640      650

      180      190      200      210
gi-13621430. PKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | | : || : || : : : || : : : : : : | : : : |
gi-19224134. PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQNNKV
      660      670      680      690

```

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:45:36 2004

! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:00.3

! Total CPU time: 0:00:00.4

! Output File: gi-13621430.fasta

FIGURE 59G

TYPE 3 pilus motifs
protein F2 like fibronectin-binding protein
 Length: 696-733
 LPXTG
 pilin motif consensus PK (X₇) K
 E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

TYPE 4 pilus motifs
protein F2 like fibronectin-binding protein
 Length: 1161
 LPXTG
 pilin motif consensus PK (X₇₋₈) K
 E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYQQVTEK
		156 PKMSVVSKYK
		674 PKYDAKNQEVK
	E boxes	563 YDLYETKAPKGY
		940 YTFVETAAPDGY

FIGURE 60

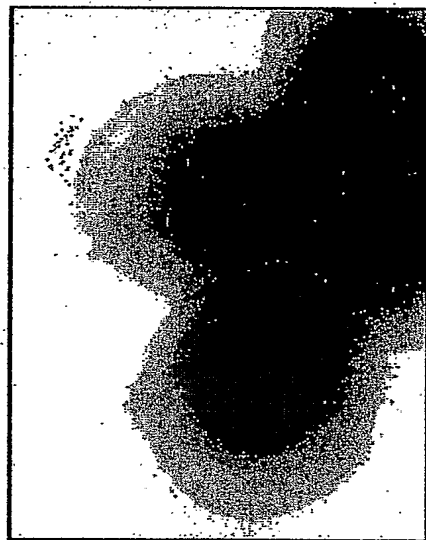
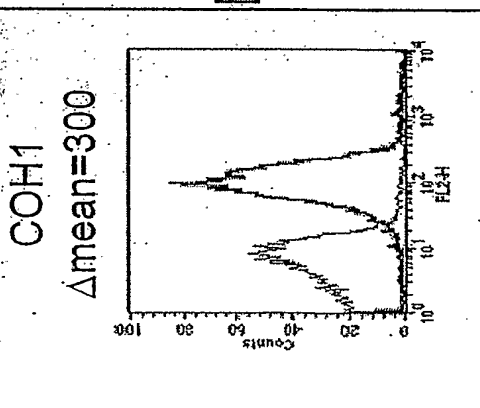
Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

WO 2006/078318

PCT/US2005/027239

211/487

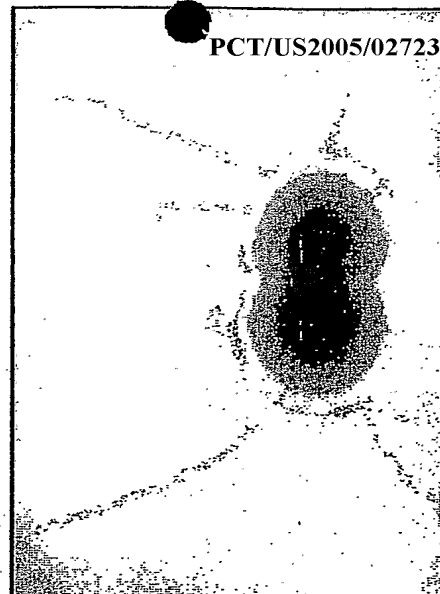
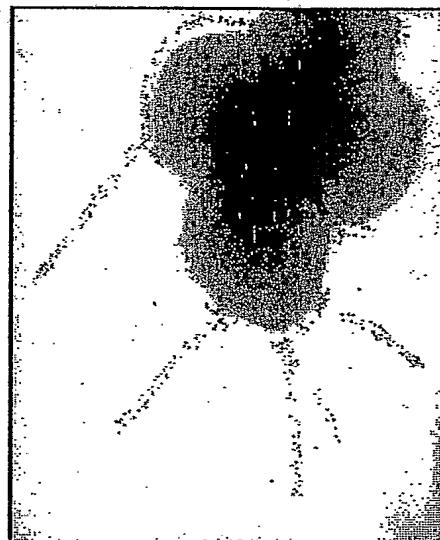
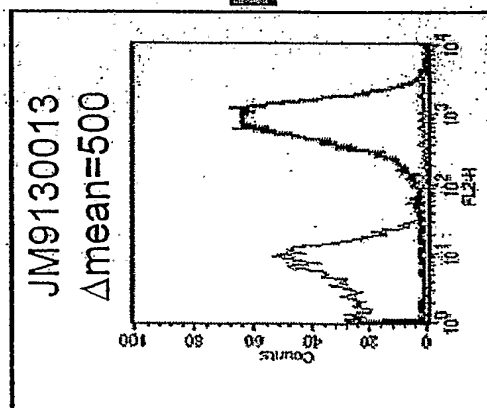
α -GBS80



Immunogold Electron Microscopy



α GBS80



PCT/US2005/027239

CHIRON VACCINES

FIGURE 61

Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

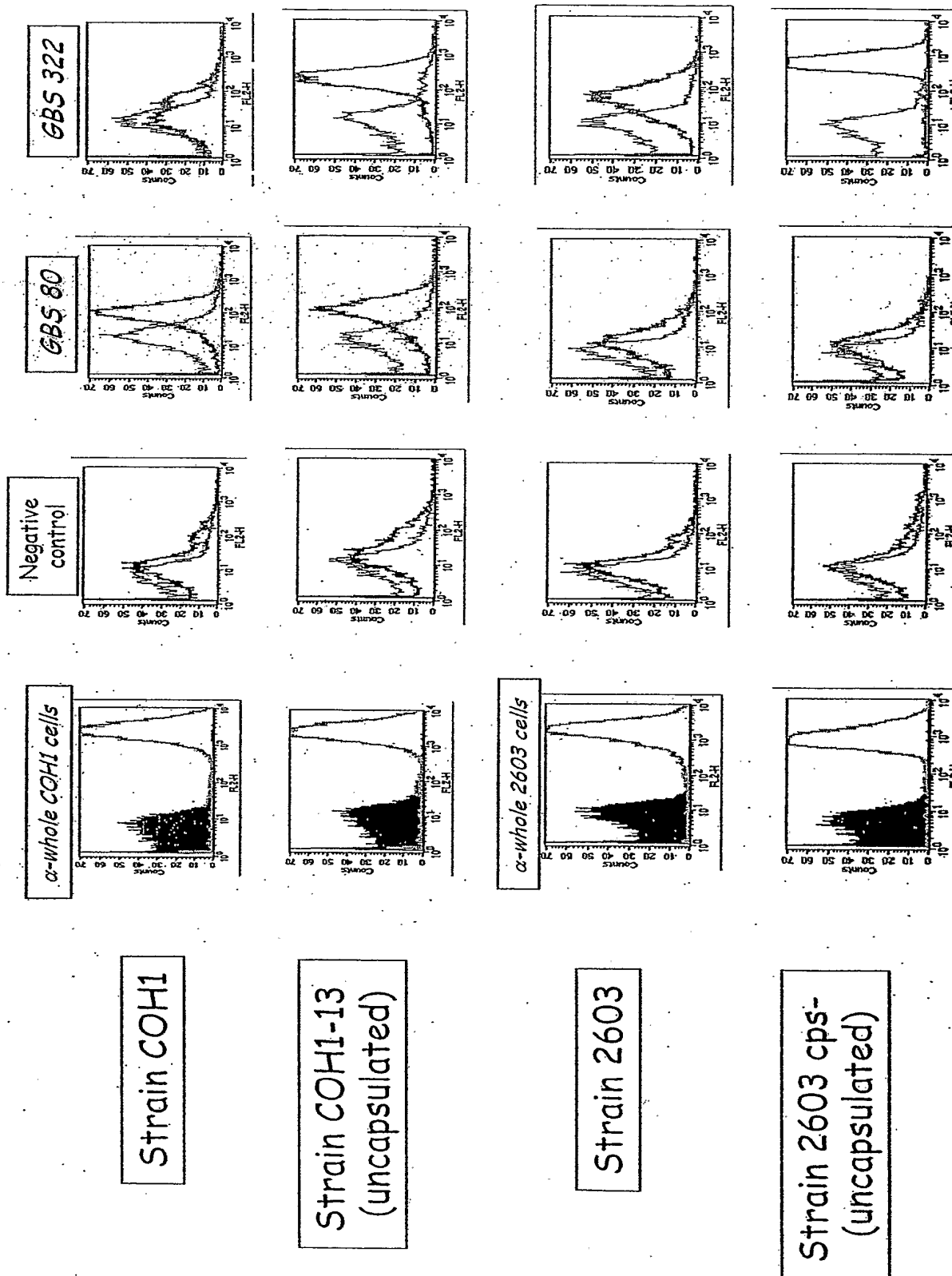
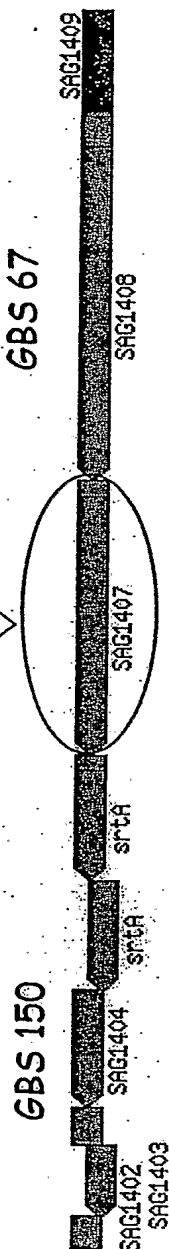


FIGURE 62

Adhesin island 2-

Operon gbs 67, 59, 150



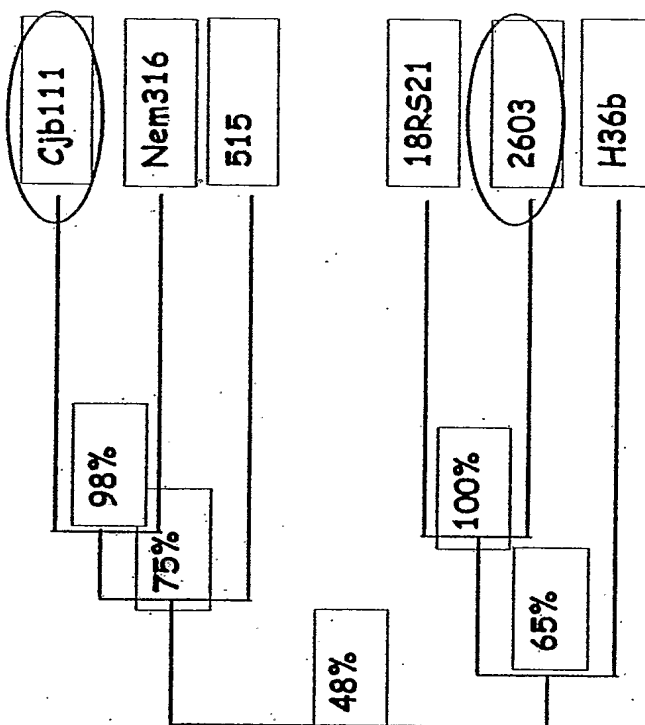
• 30 % identity with GBS 80

• By comparing GBS 59 amino acid sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

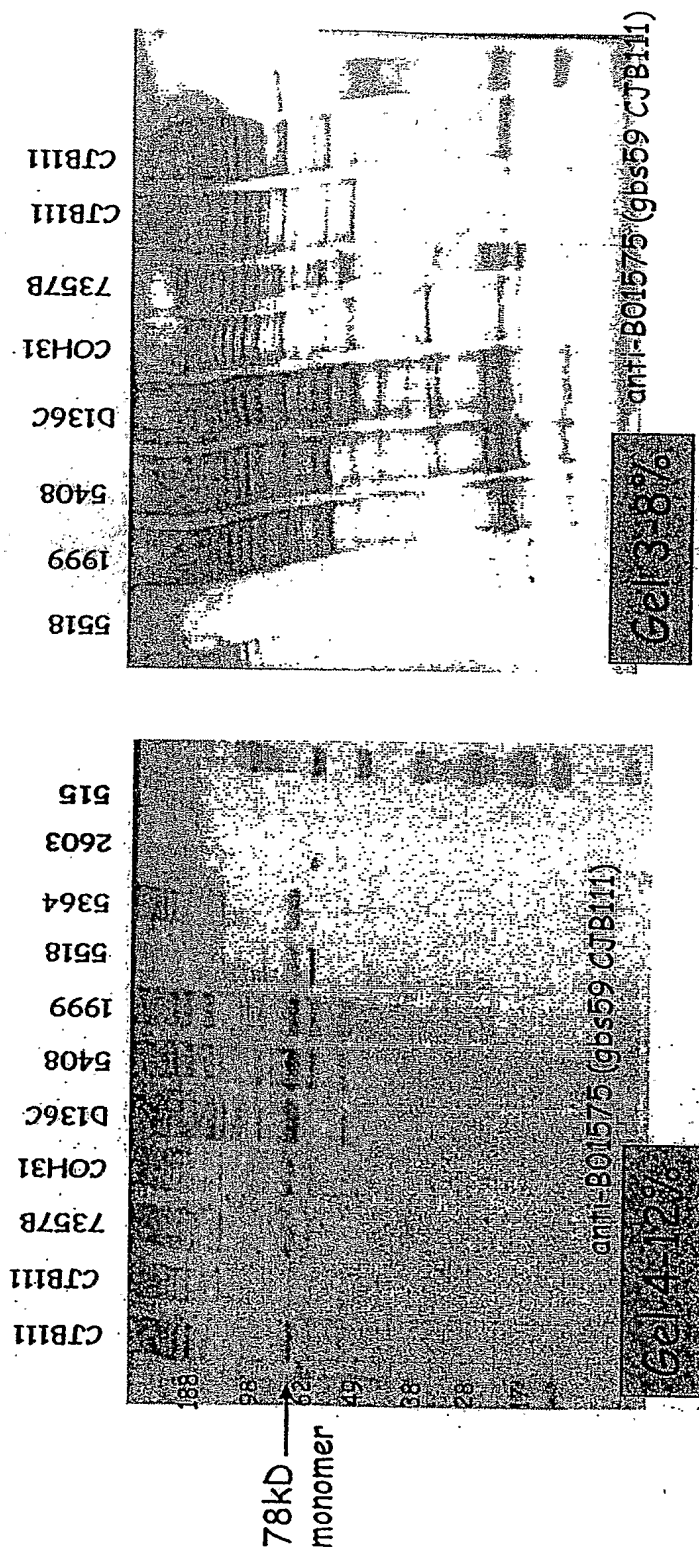
Two-by-two amino acid sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)
in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,
5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

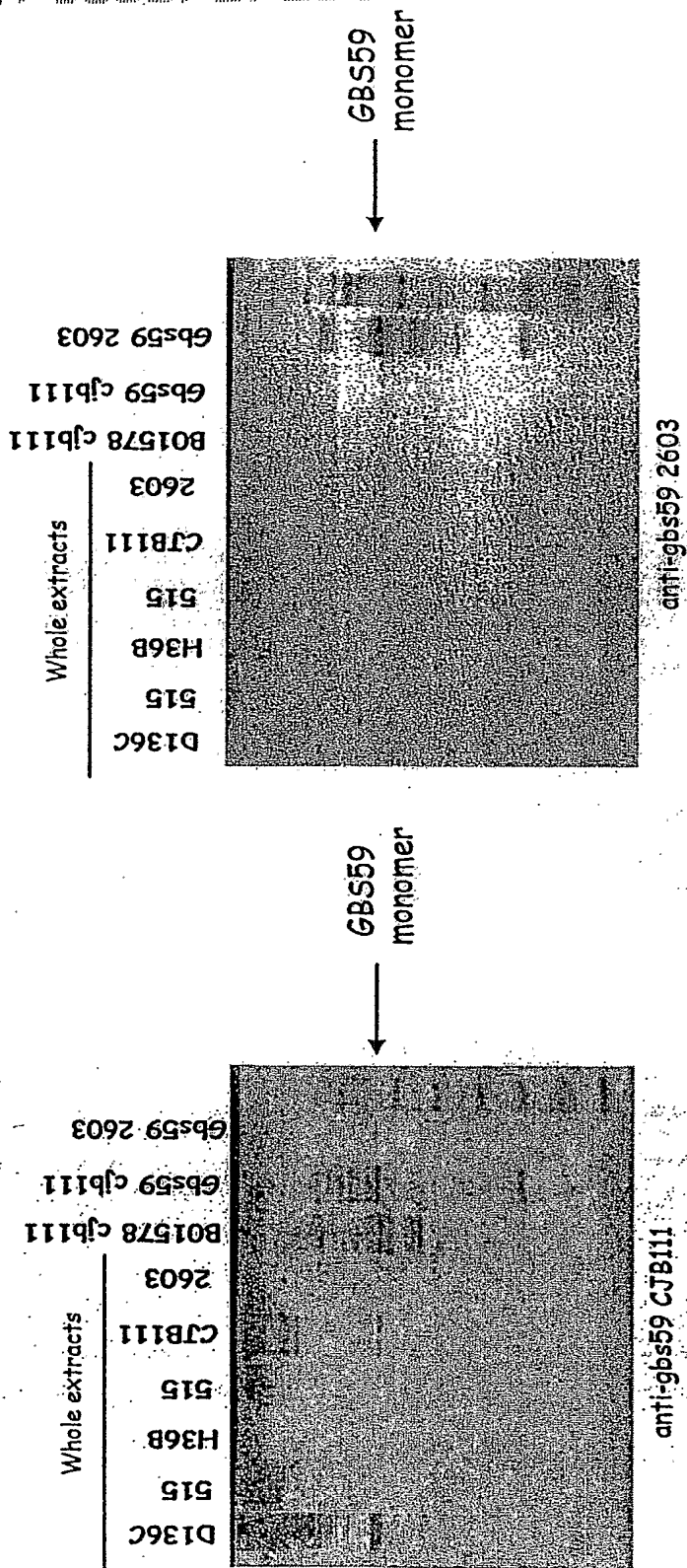


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

WO 2006/078318

216/487

PCT/US2005/027239

PCT/US05/27239

GBS strains	Type	GBS 59
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C		504
COH31	III	505
DK21	II	249
CJB111	V	493
5364		593
2110		590
1999		594
2210		636
5408		537
1169		227

GBS 59

GBS 67

PBS

CJB111

7357B

515

Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66

Opsonophagocytosis assays: B01575 (gbs59-cjb111)

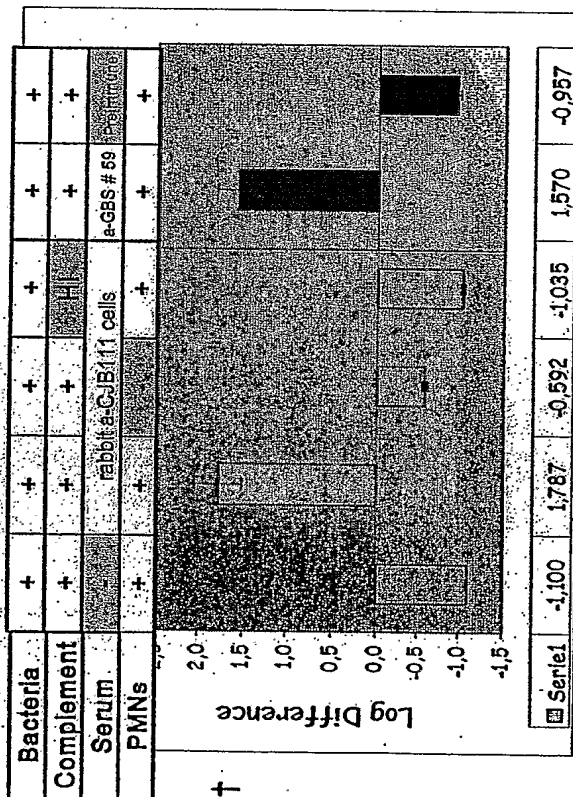
WO 2006/078318

217/487

PCT/US05/27239

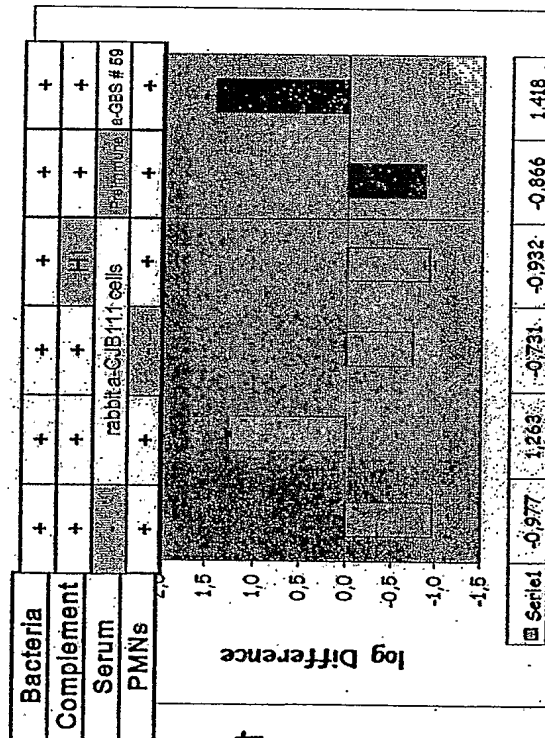
PCT/US2005/027239

I
experiment



- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

II
experiment



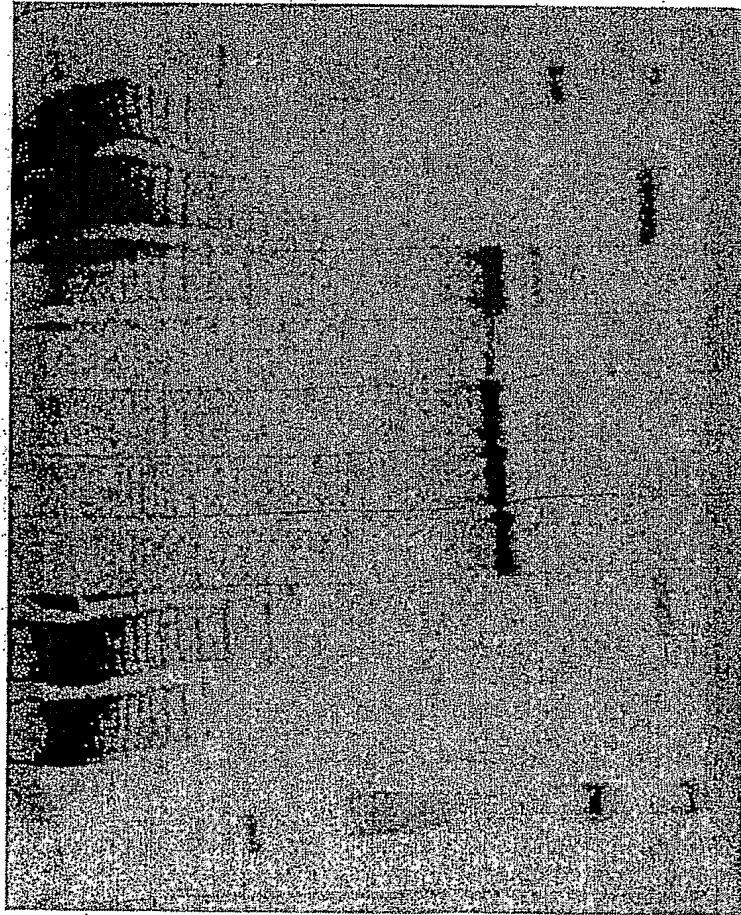
Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V

FIGURE 67

Association GBS 80-104

WB α -80, α -104 JM9130013 Total Extract

Mab α -80 A4/77
 Mab α -80 19G4/78
 Mab α -104 15H3/49
 Mab α -104 12A7/67
 Mab α -104 H2/32
 Mab α -104 14F3/73
 α -104 POLIC.
 α -80 POLIC.
 Mab α -80 19E6/77

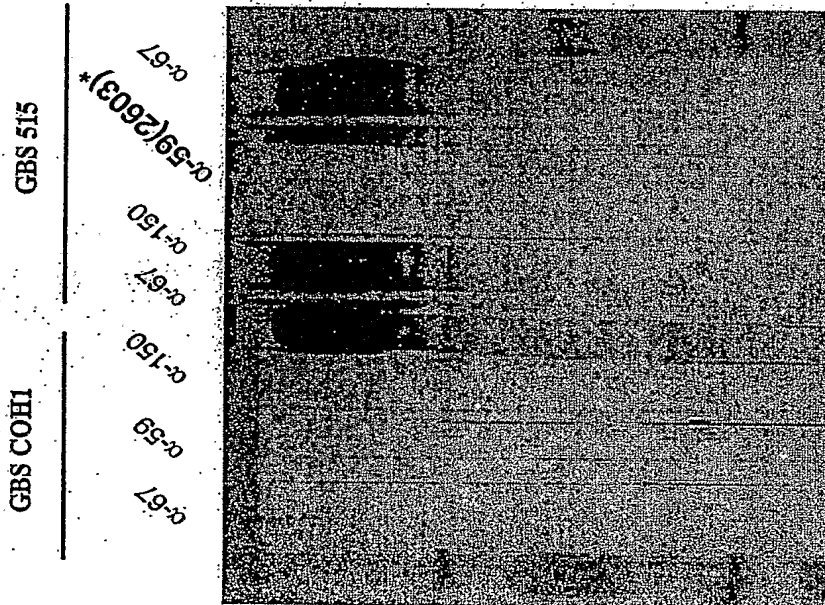


FACS	(Δ MEAN)
GBS 80	597
GBS 104	446

FIGURE 68

WB GBS 515 Total Extract

α -67; α -150; α -59 (2603)



Controls:
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

FIGURE 69

Western Blotting ko GBS67 from 515 genome (clone 1.45)

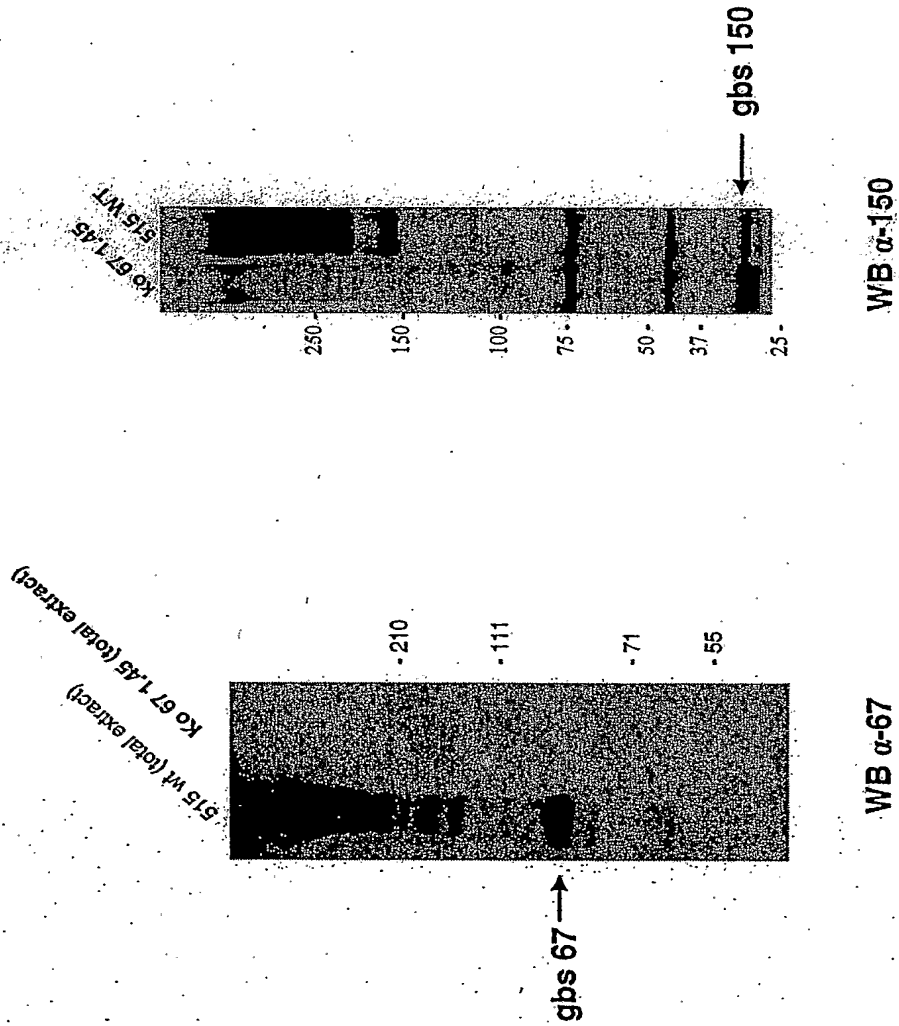


FIGURE 70

FACS GBS 515 Δ67

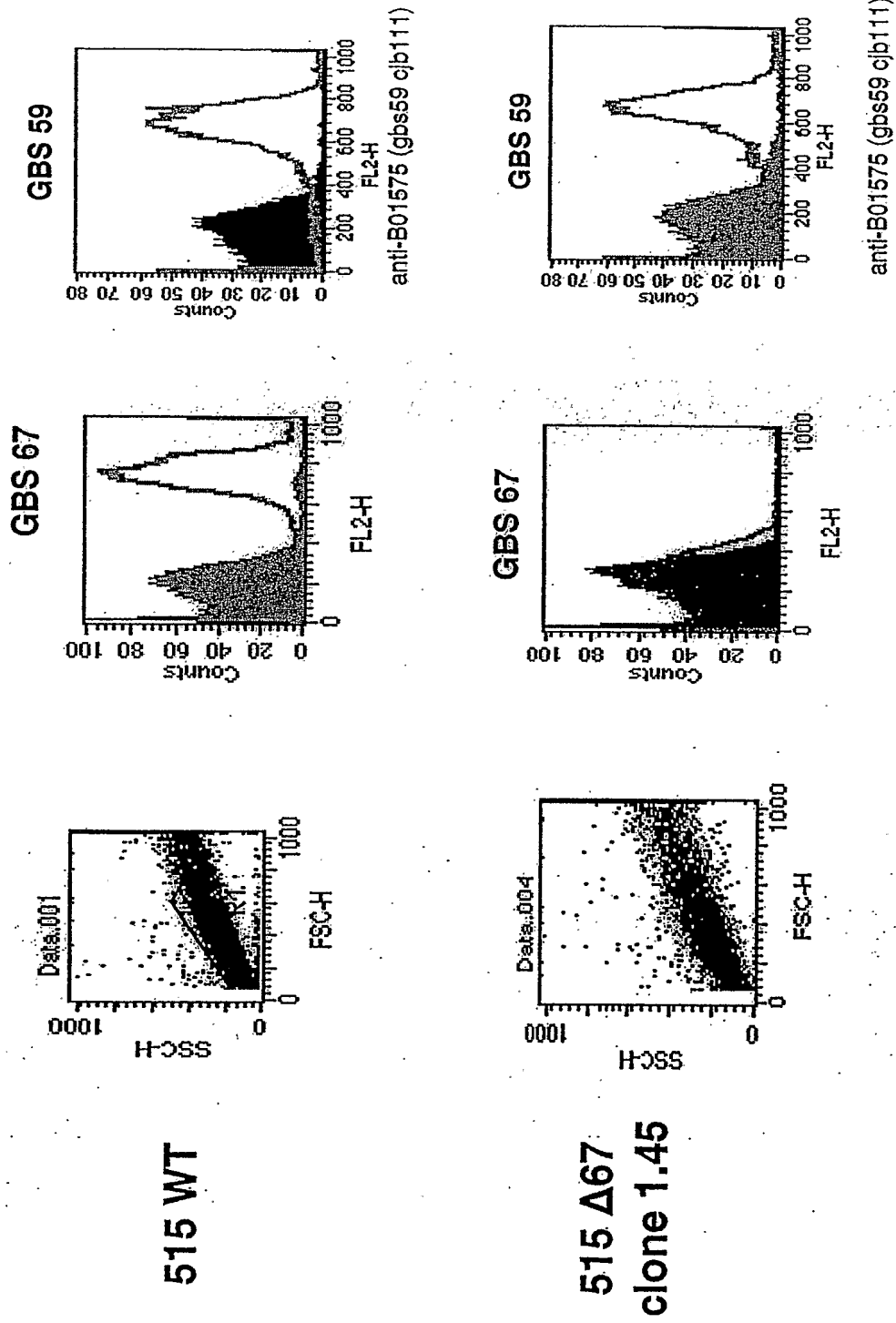


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

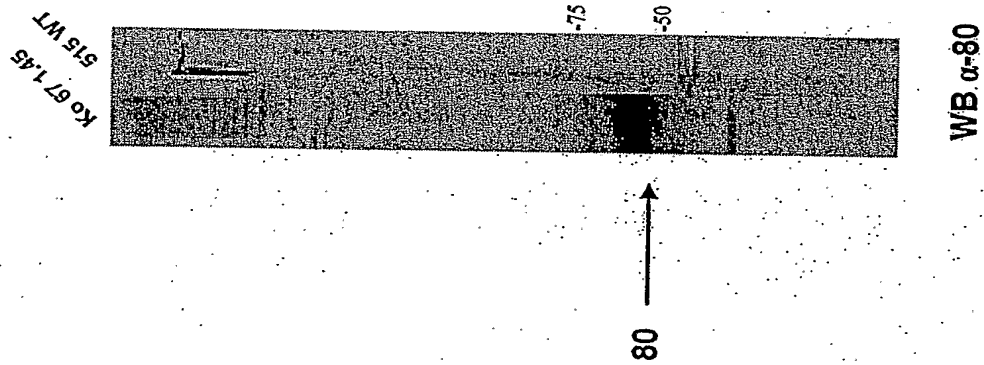


FIGURE 72

spyM6_0159 type 1 pilus present in M6

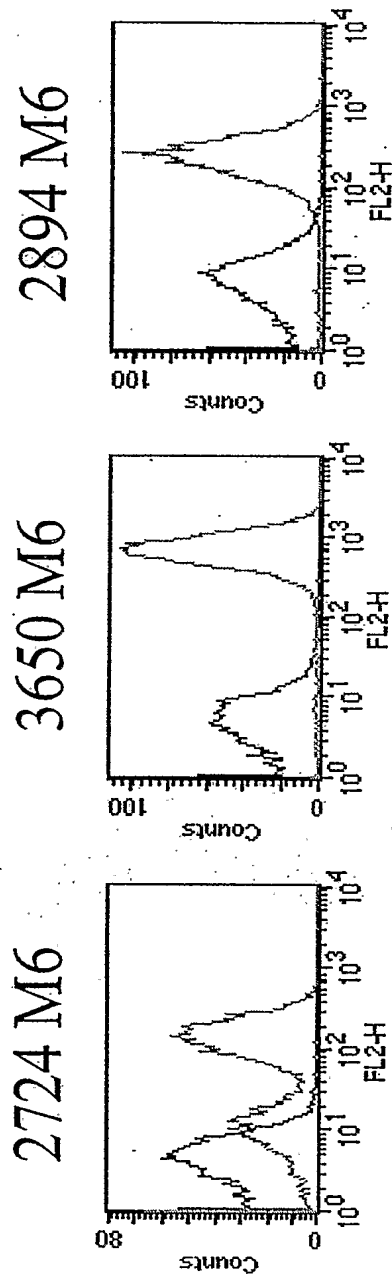


Figure 73

PCT/US05/27239 224/487

spyM6_0160 type 1 pilus present in M6

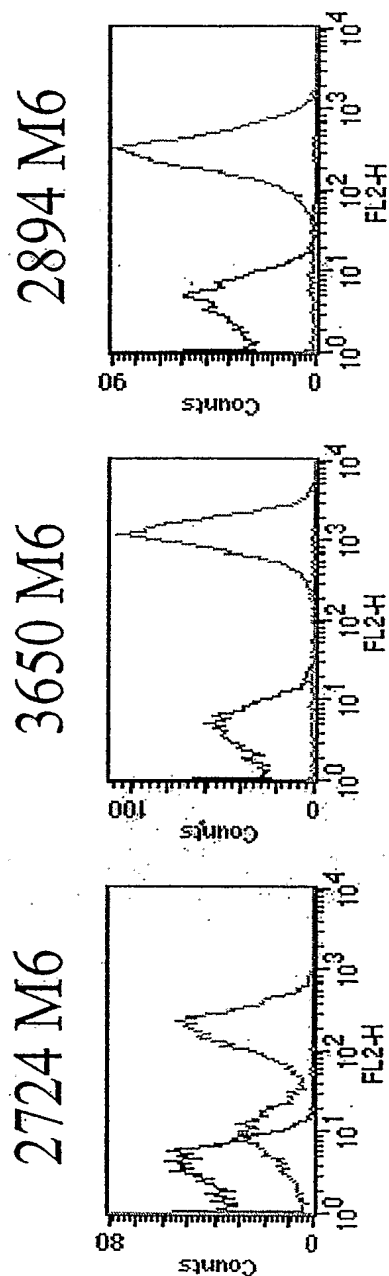


Figure 74

Gas15 type 2 pilus present in M1

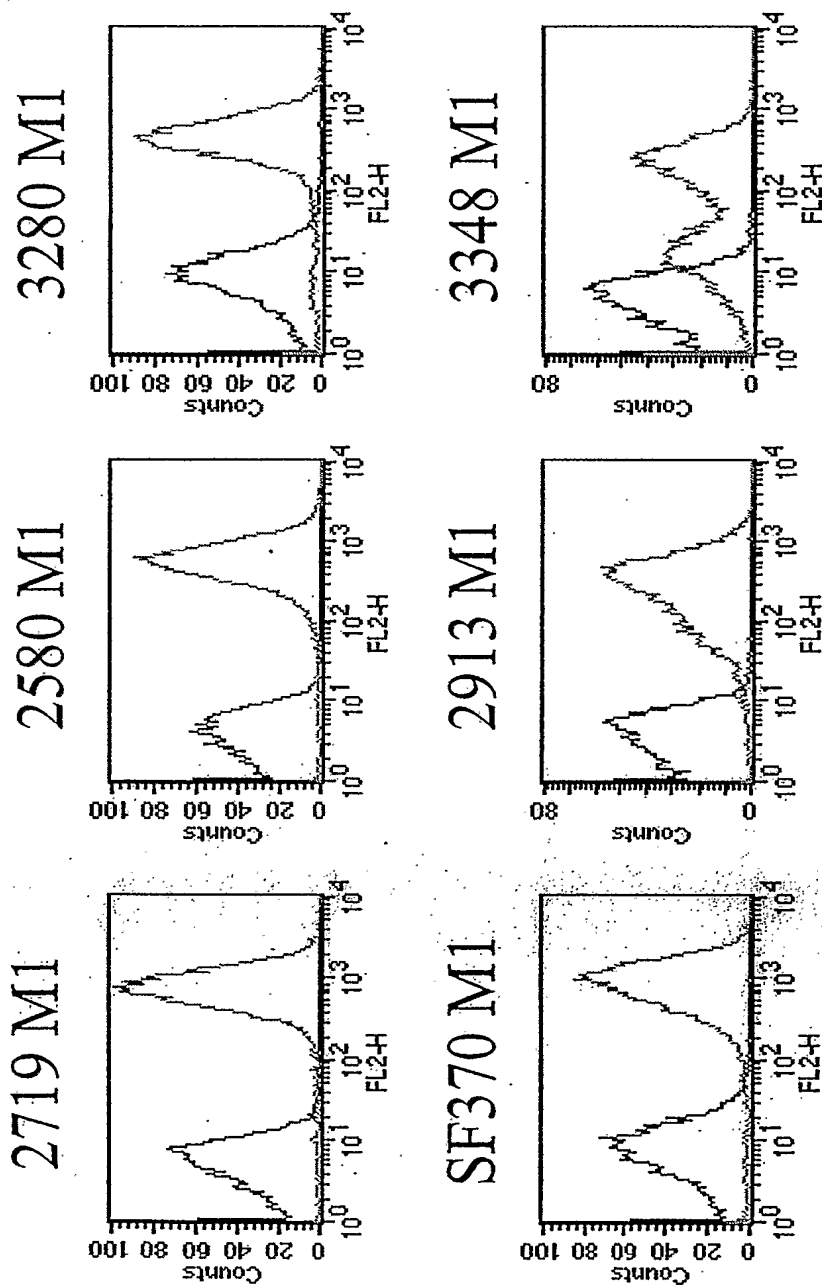


Figure 75

Gas16 type 2 pilus present in M1

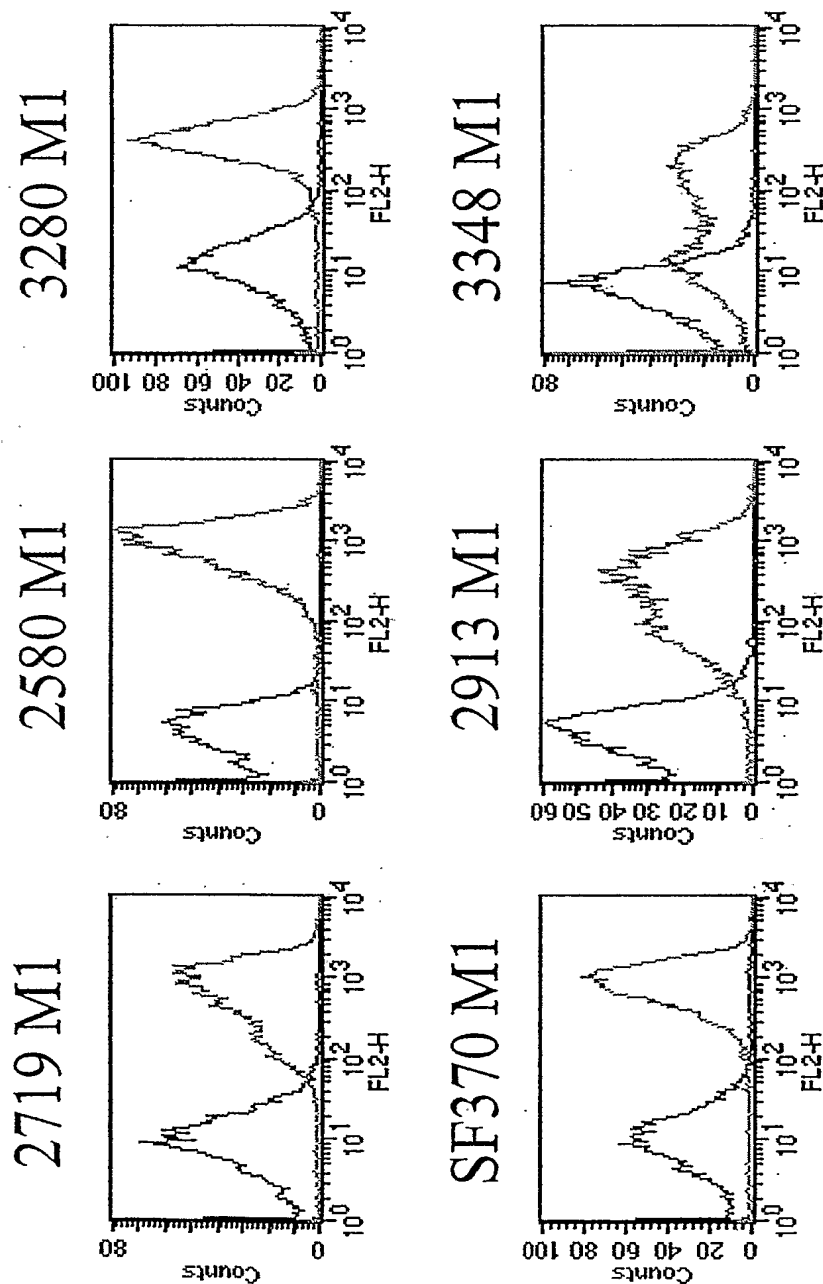


Figure 76

Gas18 serum 1 type 2 pilus present in M1

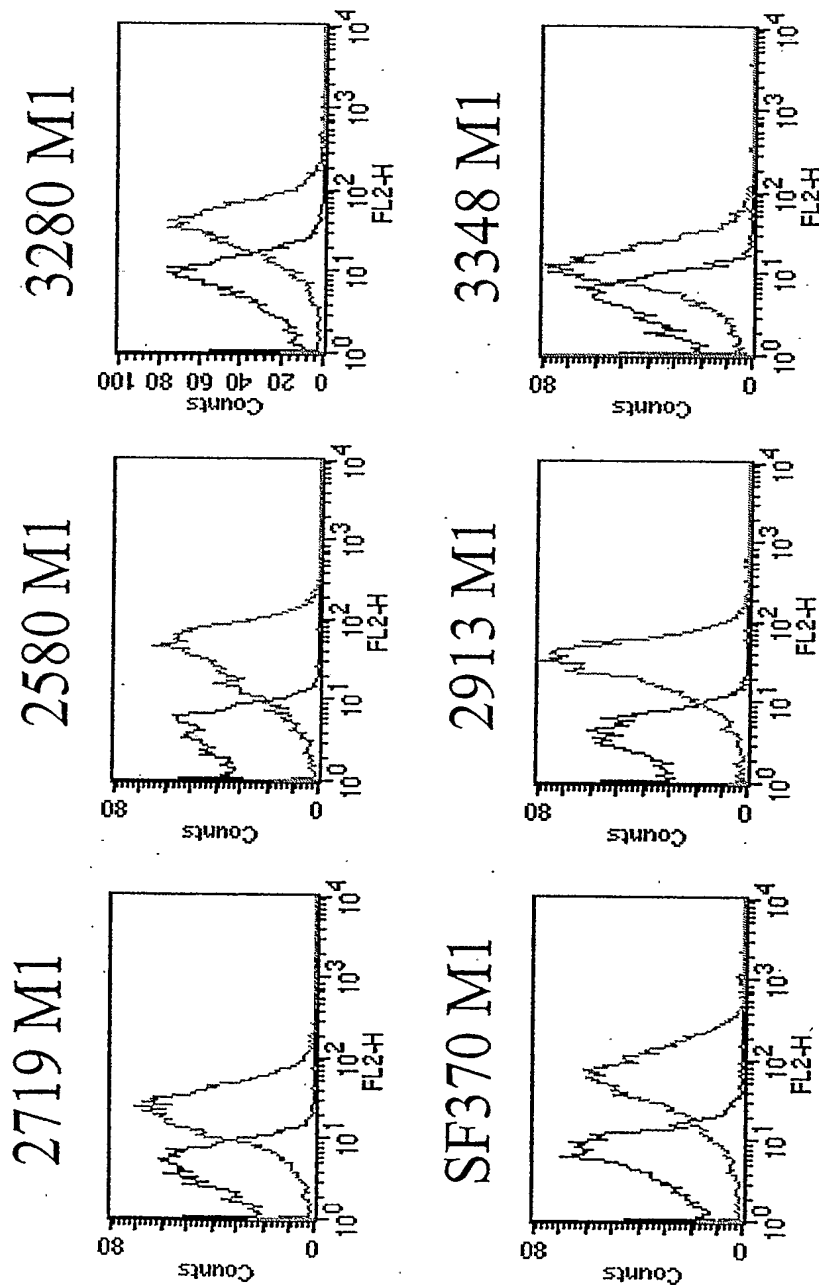


Figure 77

Gas18 serum 2 type 2 pilus present in M1

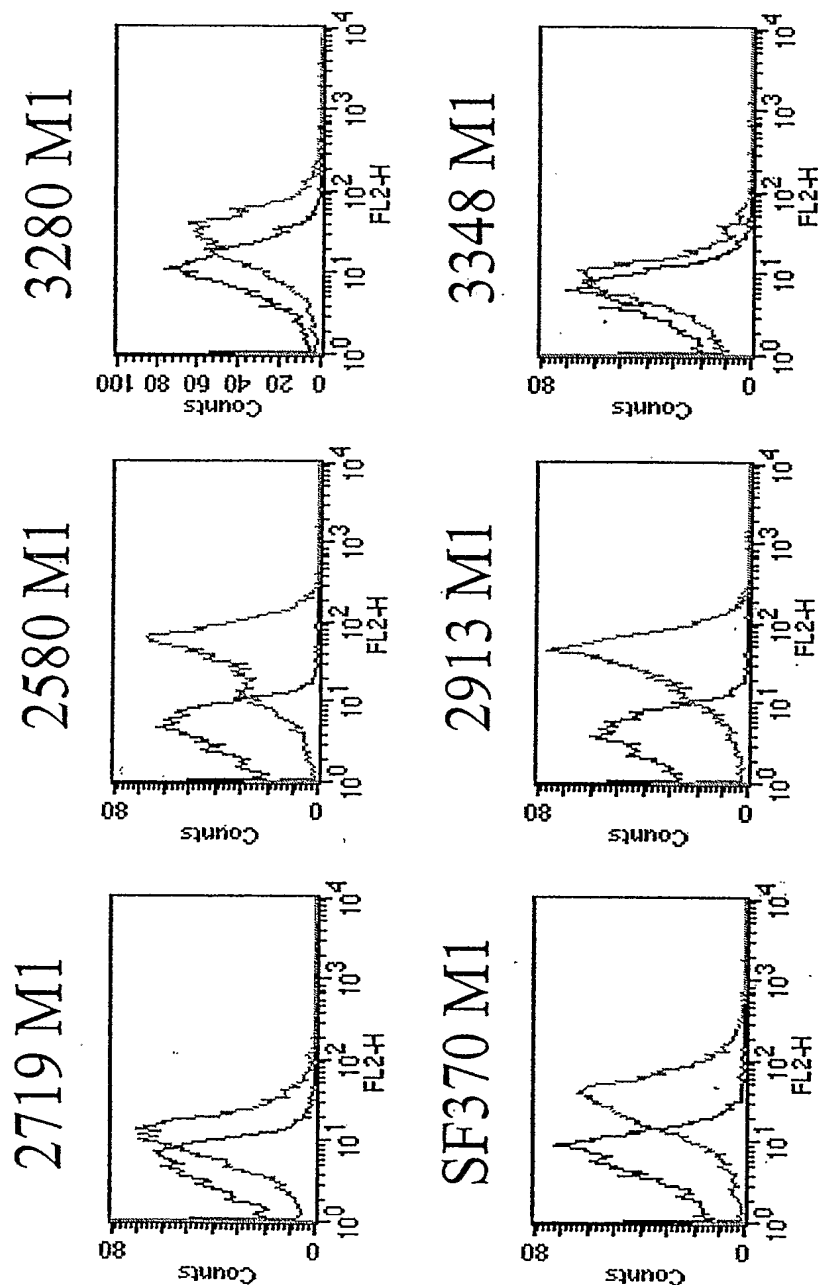


Figure 78

Gas16p2 type 2 pilus present in M1

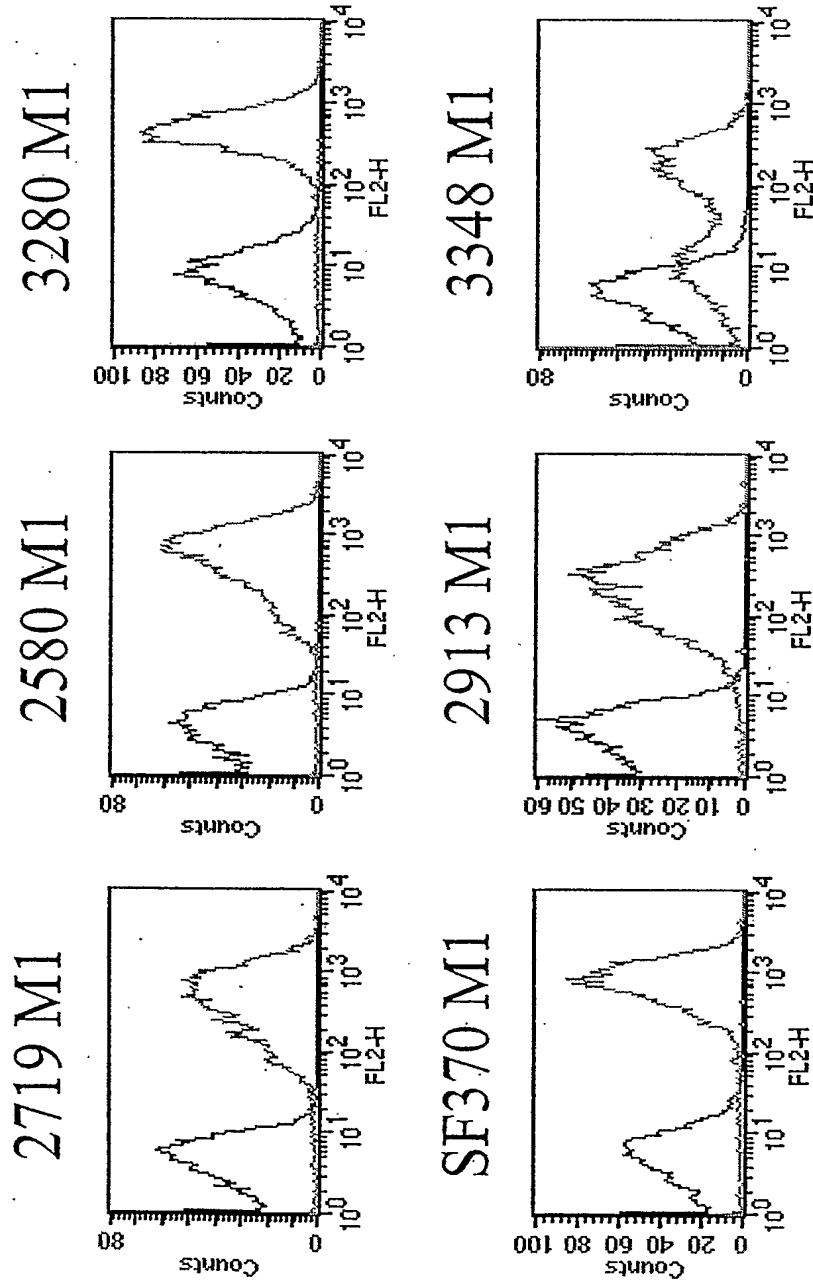


Figure 79

spyM3_0098 type 3 pilus present in M3

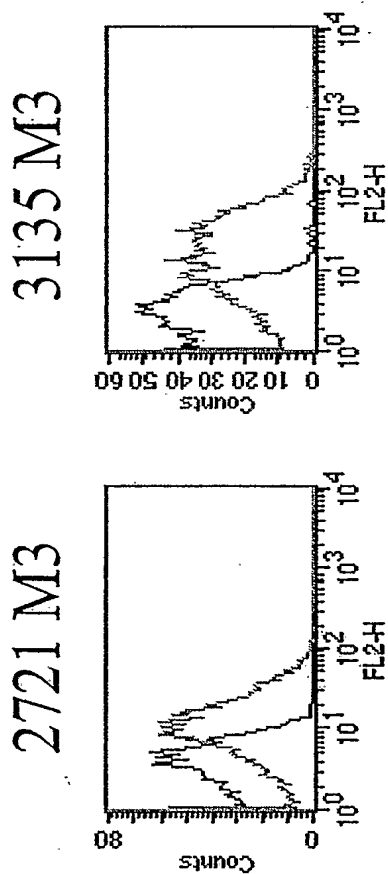
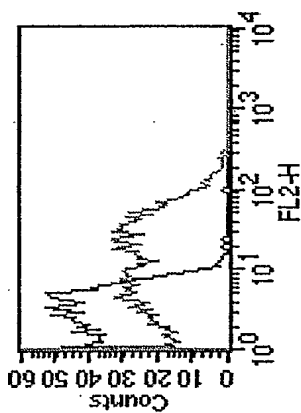


Figure 80

spyM3_0100 type 3 pilus present in M3

3135 M3



2721 M3

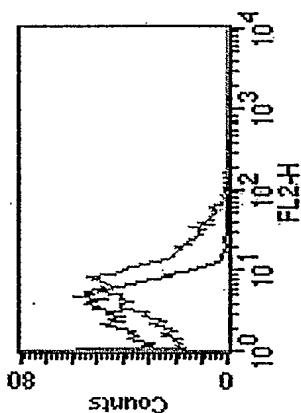
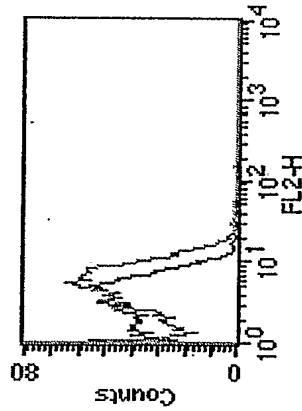


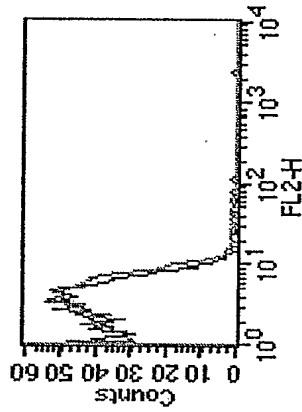
Figure 81

spyM3_0102 type 3 pilus present in M3

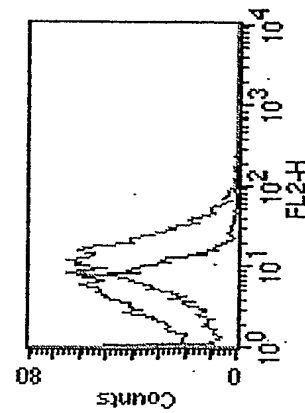
2721 M3



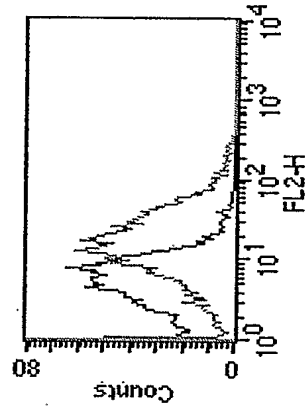
3135 M3



2724 M6



3650 M6



2894 M6

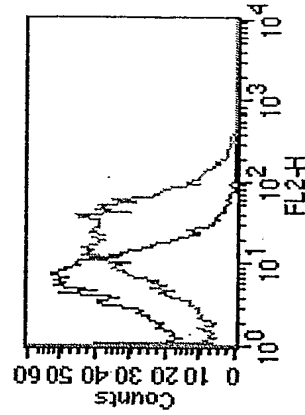
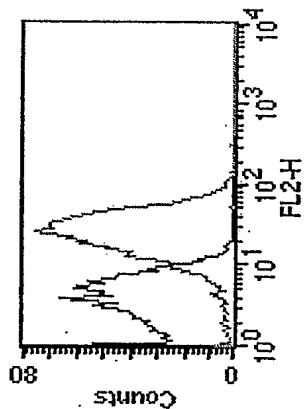


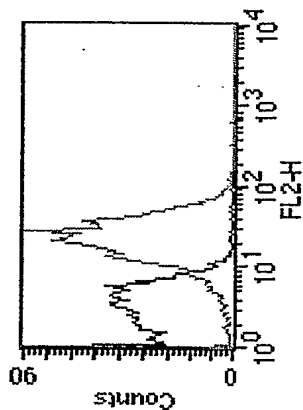
Figure 82

spyM3_0104 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

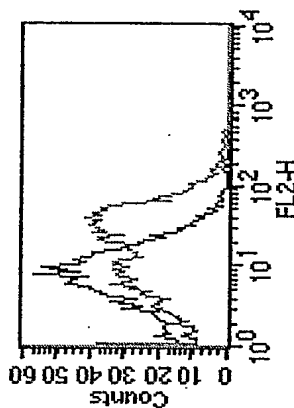
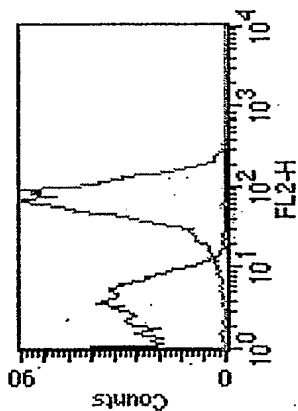


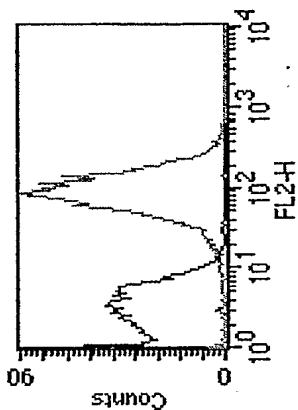
Figure 83

spyM3_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

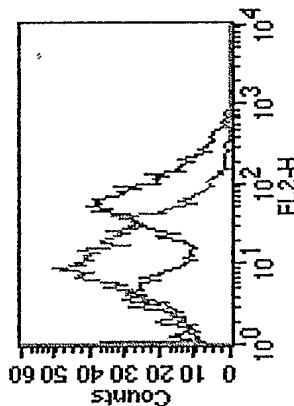
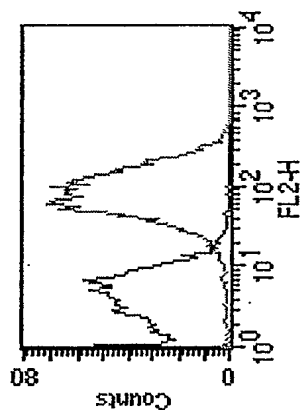


Figure 84

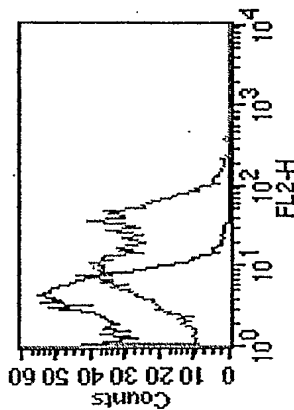
PCT/US05/27239 235/487

19224134 type 4 pilus present in M12

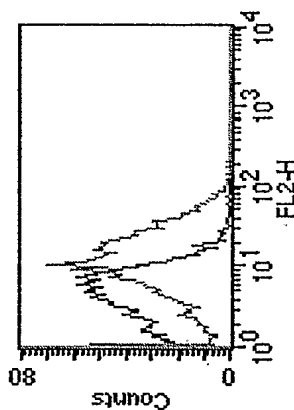
2728 M12



2724 M6



3650 M6



2894 M6

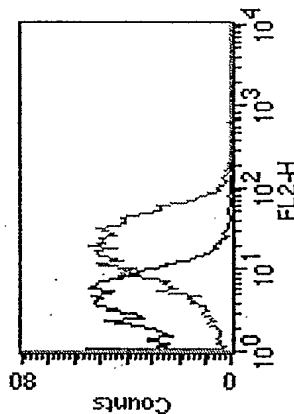


Figure 85

19224135 type 4 pilus present in M12

2728 M12

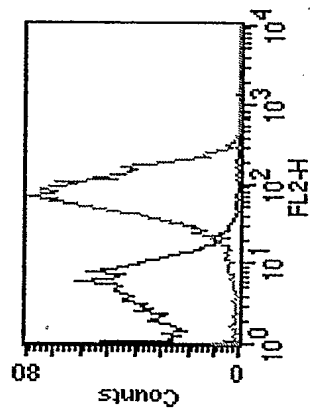


Figure 86

19224137 type 4 pilus present in M12

2728 M12

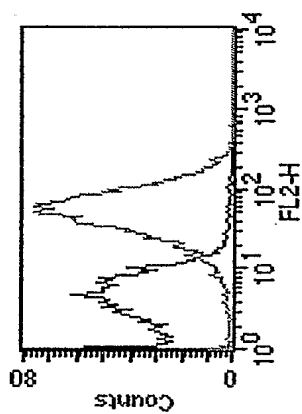


Figure 87

19224141 type 4 pilus present in M12

2728 M12

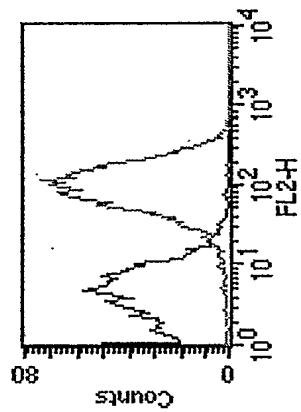


Figure 88

Figure 89

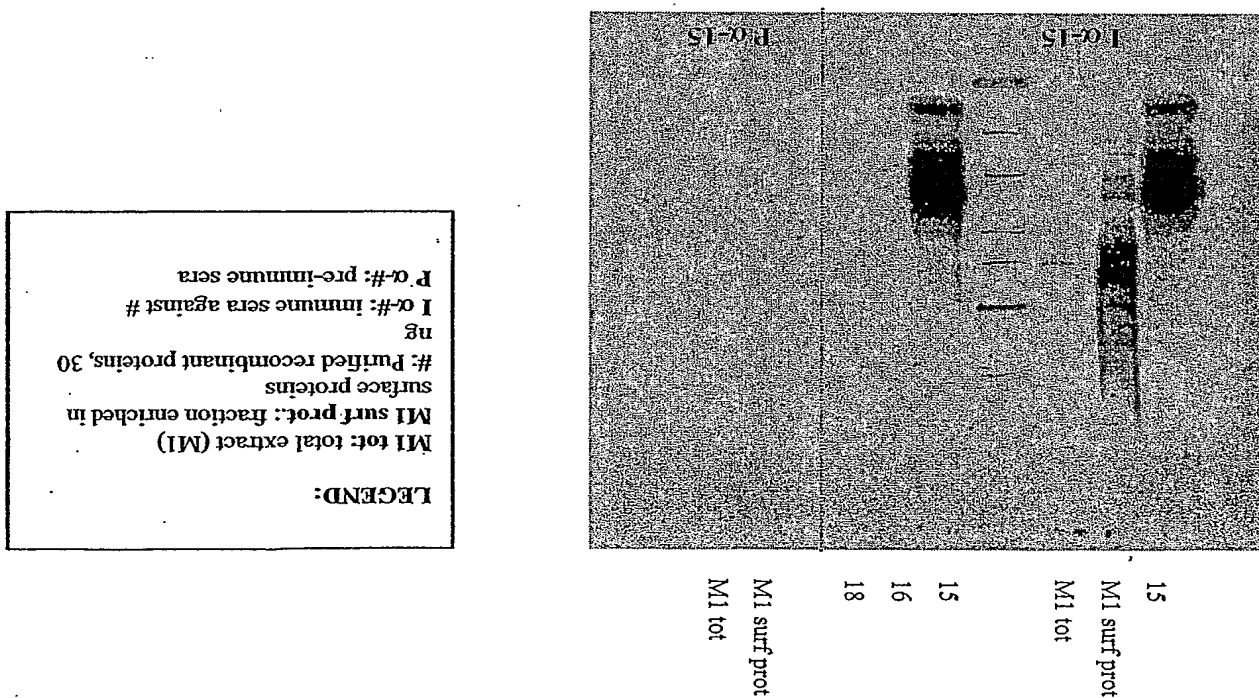
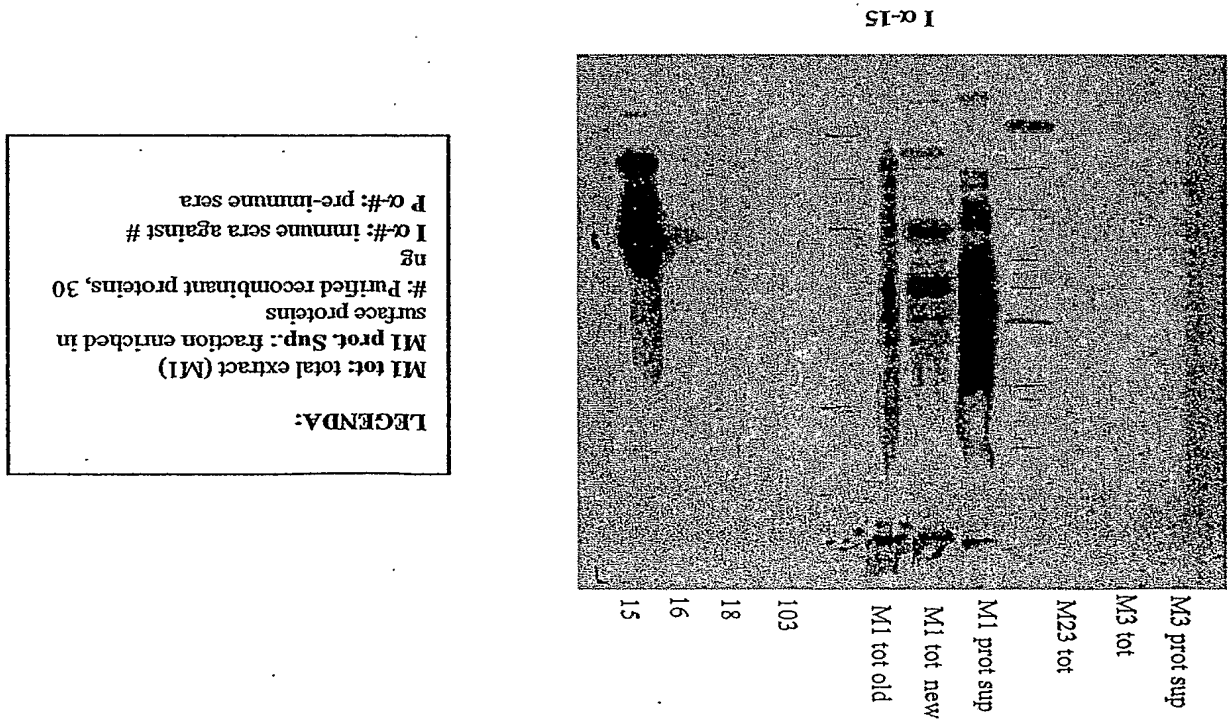


Figure 90



PCT/US05/27239 341/487

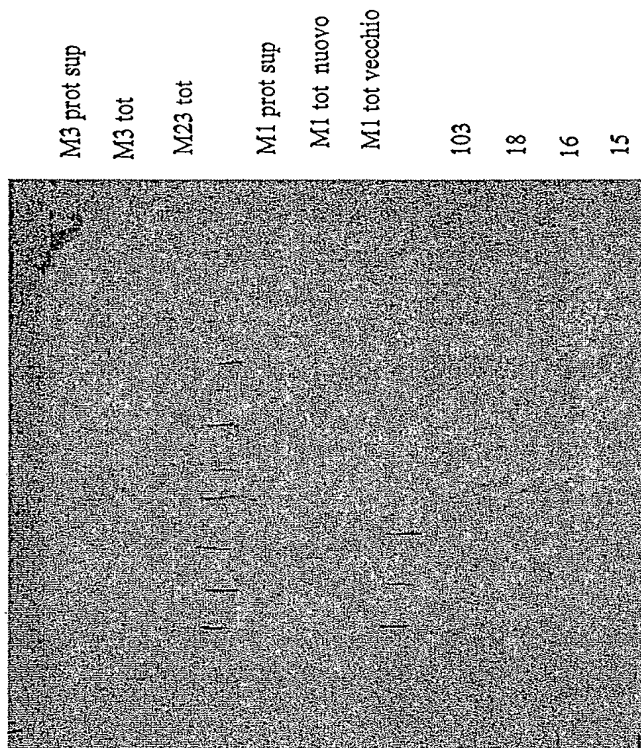
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-15**

Figure 91

PCT/US05/27239

242/487

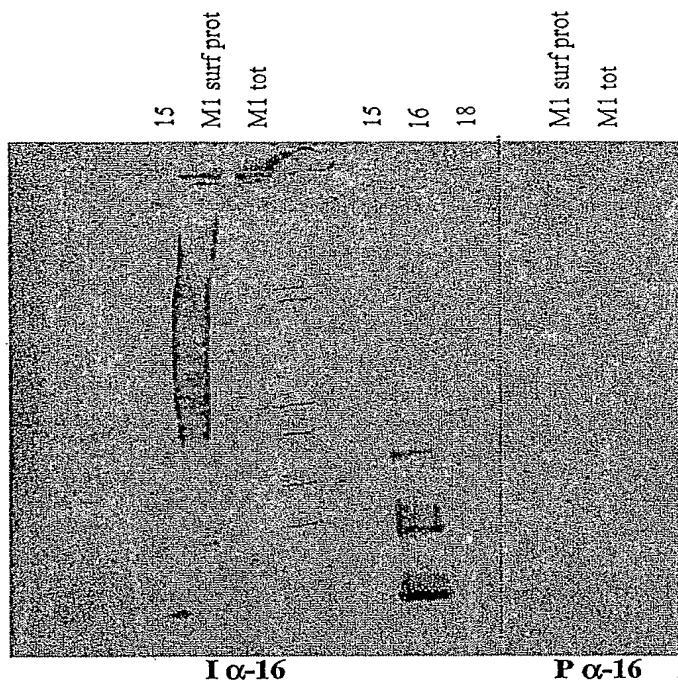
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 92

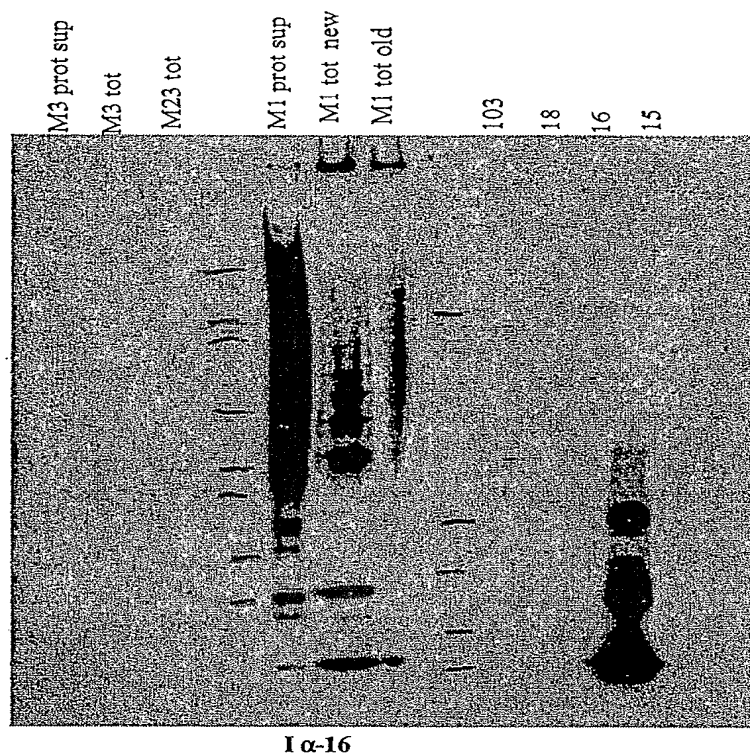
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 93

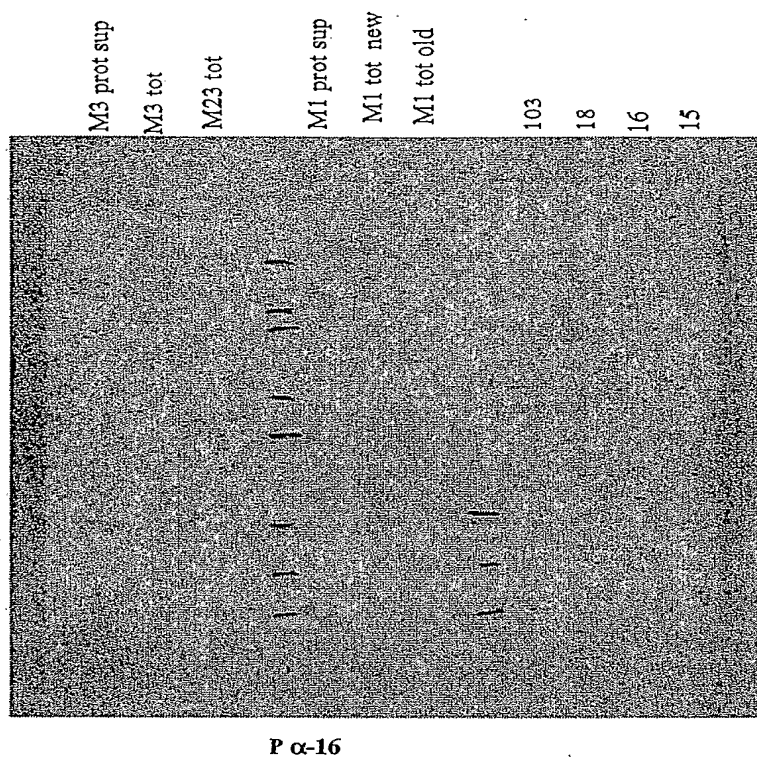
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 94

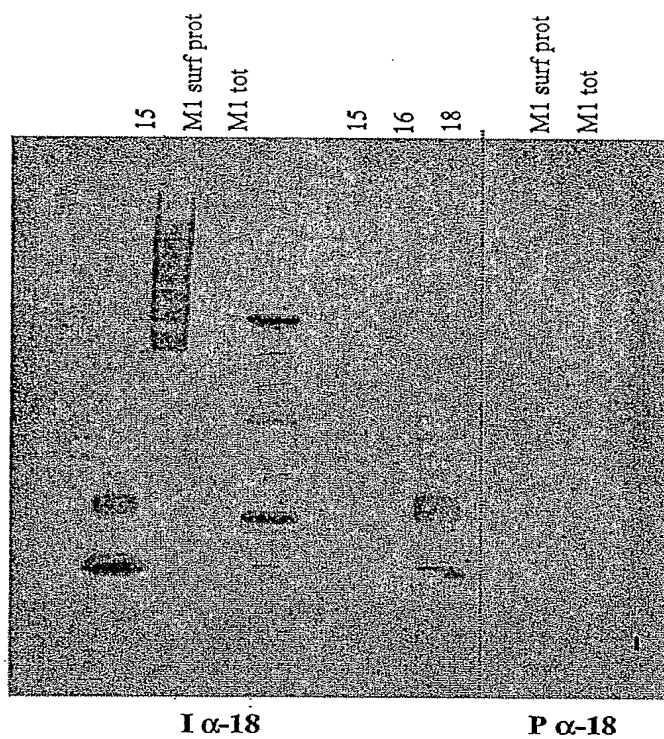
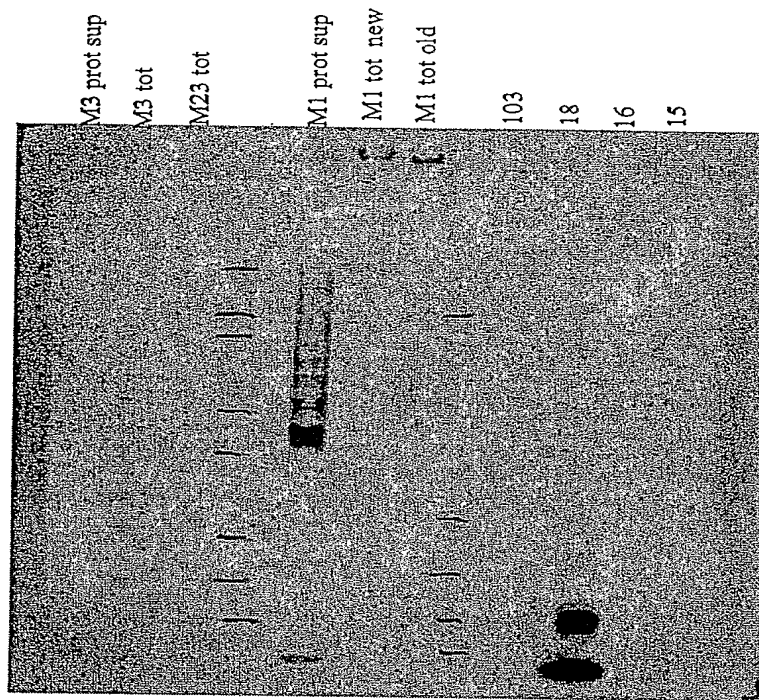
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 95

PCT/US05/27239 246/487

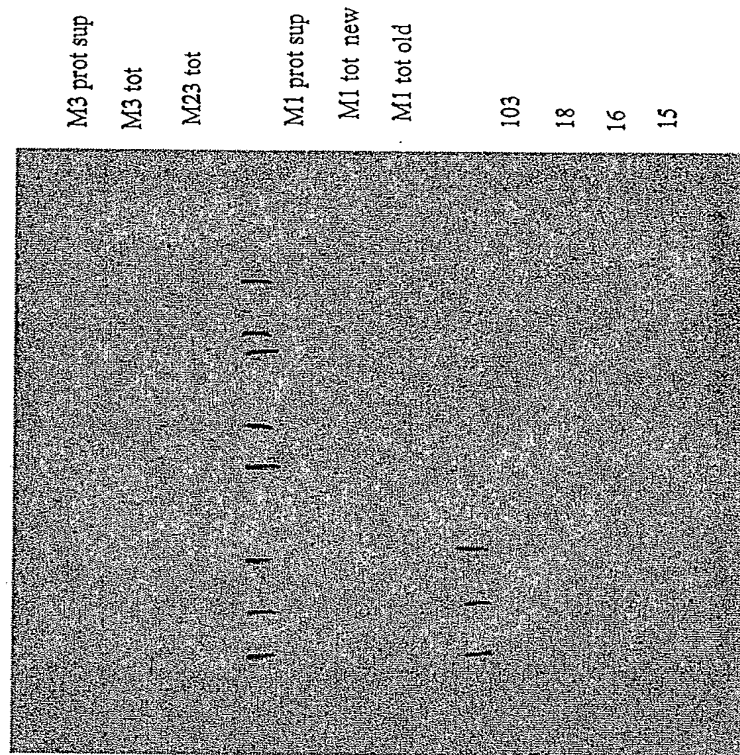


I α-18

LEGENDA:**M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 96

PCT/US05/27239 247/487

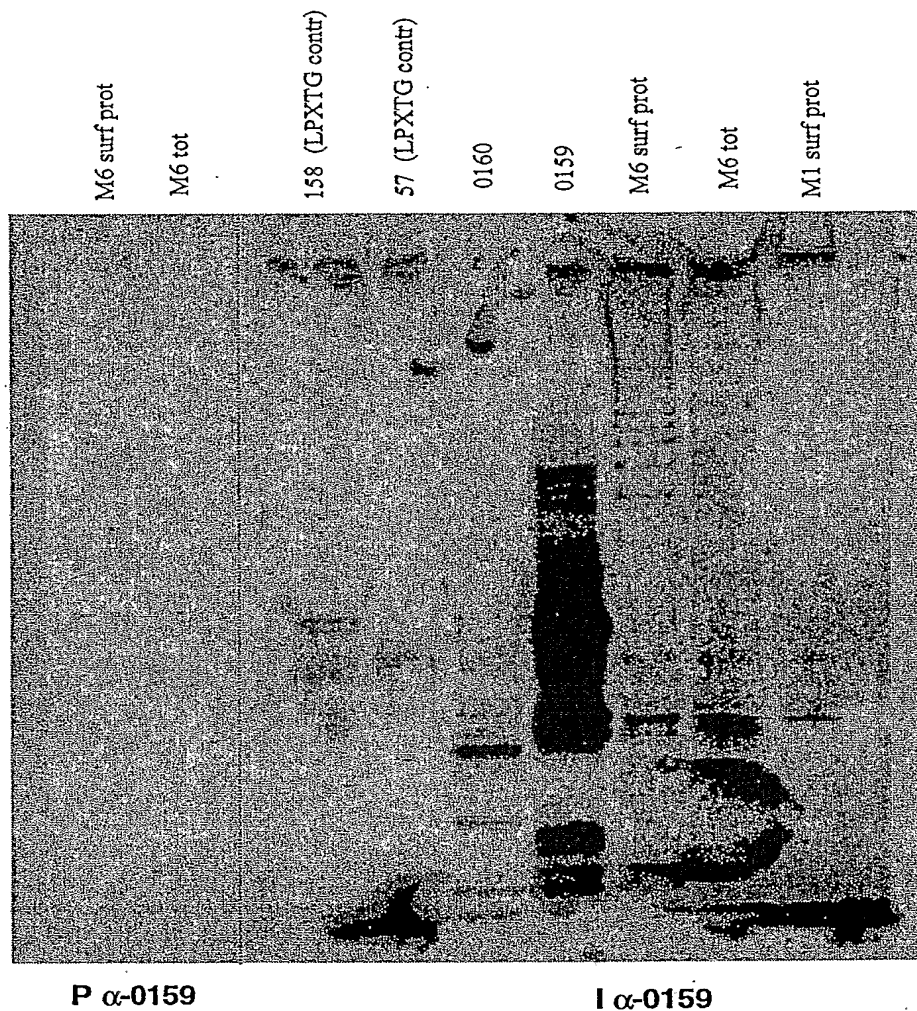
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

P α-18

Figure 97

PCT/US05/27239 248/487

Figure 98

**LEGEND:****M6 tot:** total extract (M6)**M6 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

PCT/US2005/027239 249/487

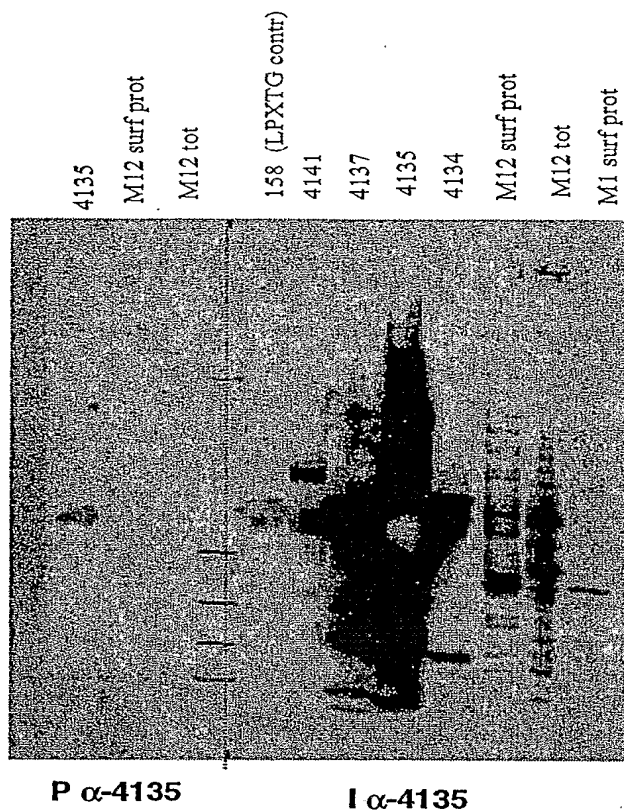
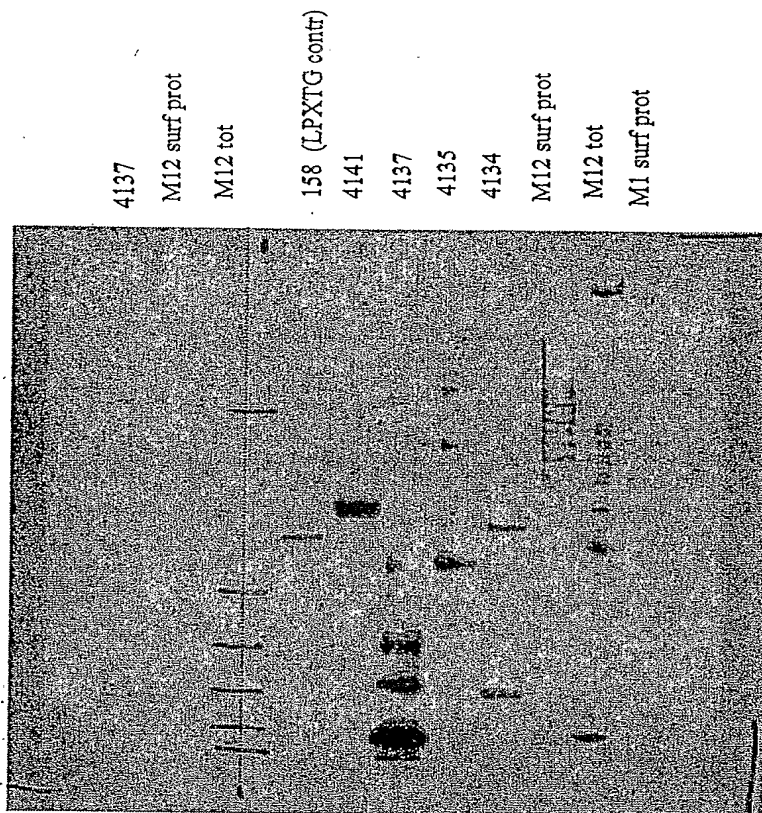
**LEGEND:****M12 tot:** total extract (M12)**M12 surf prot:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 99

PCT/US05/27239 250/487

**LEGEND:**

M12 tot: total extract (M12)

M12 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

P α-4137

I α-4137

Figure 100

FIGURE 101

1 GACAGCTTCCTTATACGACCGCTTCTATATCGGACCTTCCAGTTCTTGCTCTTTTA 60
CTGTTGGAAGGAATATGCTGGCGAAAGATATAGCCTGAGAAGTTCAAGAACGAGAAAAT
61 CCAGGACTAGCCGTTCAAGTGCACGATTGTGGTGGTTGTTAAATGGAACCAAGTCGT
GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAACAACAAATTACCTTGGTCAGCA 120
121 TCAGTTGTGACAGAAATTTACCTTCTATCTTGGGATTCCTCGTTATGTTGGAGCTAGTGCC
AGTCAACACTGTCTTAAATGGAAGATAGAACCTTAAGGCAATACAAACCTCGATCACGG 180
181 TTAAGATTTTCAAAATTTGTGAAAGCCGGAGAACTCTTGAGCTTTGGGCAATGTTTTTG
AATTCTAAAAGTTTAAACACTTTCGGCCTCTTGAGAACTCGAAACCCGTTAACAATAAC 240
241 CTCCTGGTCGCGATGGGAGTAGCTTTTGGCGTCAGATGGTGGCTATTCTGCTCTTGACC
GAGAACCAAGCGCTACCTTCATCGAAACAGCCAGTCGTACCAACCGATAAGCGAAGACTGG 300
301 AGCTATGTGAAAAACACGACTTCACCCCTTTTGGTAAATAACCGTATCGTGTGGTAGT
TCGATACACTTTTGTGCTGAAGTGGGAAAAACCATTTATGGCATAGCACGAACCATCA 360
361 GTTTTGCTACTTTACAGTTTGTCCGTTTATTTGTATAAGAAAAACCTTGAGGGGTAAAC
CAAAACGATGAATGTCAAAACAGGCAATAAACAATATCTTTTGGAACTTCCCCCATTG 420
421 TCTTCAAGGTTTTTATACCTTAGAAATCTCTTCAAAACCGGCTCAGCTTTATCTGCAACC
AGAAGTTCAAAATATGAGAATCTTTTAGAGAAGTTTGGCGGCAATAGACCGTTGG 480

Figure 101A

481 TCATAAACAGTGTGTTTGAGCAGCCCTGCGGCTAGCTTCCTAGTTTCTCTCTTCGATTTTCATT
-----+-----+-----+-----+-----+-----+-----+-----+
AGTTTGTGCACAAAACCTCGTCGGACGCGGATCGAAGGATCAACGAGAACTAAAGTAA 540
-----+-----+-----+-----+-----+-----+-----+-----+
GAGCTTTAAATCCAGTCAGGGTAATCCCCAATAGCGGACACCTCTTTCTTTCTCTCGCTT 600
-----+-----+-----+-----+-----+-----+-----+-----+
CTCGAAATTTTAGGTCAGTCCCATTTAGGGTTATTCGCCCTGTGGAGAAAGAAAGAGCGAA
-----+-----+-----+-----+-----+-----+-----+-----+
AATCTTCATAGAGTTGCAGGGCTATTTGGCTTATCTGACTAGCATCTTGTTGTTTTTGG 660
-----+-----+-----+-----+-----+-----+-----+-----+
TTAAGAAGTATCTCAACGTCGCCGATAAACCGAATAGACTGATCGTAGAACACAAAAACC
-----+-----+-----+-----+-----+-----+-----+-----+
CAAGACTTTTTCGTTTGGTAAGAGTTGAAAAAGTCCCTCGTAGCGGATTTTCAAAATGACAA 720
-----+-----+-----+-----+-----+-----+-----+-----+
GTTCTGAAAAAGCAACCATTTCTCAACTTTTCAGGACATCGCCTAAAGTTTTACTGTT
-----+-----+-----+-----+-----+-----+-----+-----+
TTTTTCCAGCTTTTCTTTGTTGATGTAGATTGAGAGGACTTTTCTGTATAGAAAGATCA 780
-----+-----+-----+-----+-----+-----+-----+-----+
AAAAAGGTCGAAAAAGAACAACTACATCTAACTCTCGCTGAAAAAGACTATCTTCTCAGT
-----+-----+-----+-----+-----+-----+-----+-----+
GCTCTTTTGTGATATCTTCTCGGCACGGAGAAATCTTCCCGTAGGTTTCTCTCTTGCCGA 840
-----+-----+-----+-----+-----+-----+-----+-----+
CGAGAAAAAACTATAGAGGAGCCGTGCTCTTTAGAGGGCATCCAAAAGAGGAACGGCT
-----+-----+-----+-----+-----+-----+-----+-----+
TTGATTACGGATCGGATTGGATTGACTGGAGAGTTGTGAATGCCACGAGCCTTTCCGAT 900
-----+-----+-----+-----+-----+-----+-----+-----+
AATAAATGCTACGCTAACCTAACTGACCTCTCAACACTTACGGTGTCTGGAAAGCTA
-----+-----+-----+-----+-----+-----+-----+-----+
ACAGATCATAGCCTAGTCTACCAAAAACGGTCTATTAGGTTACTCAGGAACCTTCAAGTA 960
-----+-----+-----+-----+-----+-----+-----+-----+
TGTCATATCGGATCAGATGGTTTTTGGCAGATAATCCCAATGGAGTCCCTTGAAGTTCTAT

253/487

Figure 101B

961 AATCAGCACCAAGTAAACAGCCCATTTGATGAAGACGTTCTACTGCTCTTTTCCCTACTC
-----+-----+-----+-----+-----+-----+-----+
TTAGTCGTGGTCATTTTTCGGGTAAACTACTTCTGCAAGATGACAGAAAAAAGGATGAG
-----+-----+-----+-----+-----+-----+-----+ 1020

1021 CATGAAATTTGGAATATCCATTTGTTTGAGAAAAATCCTCAGCCCTGTTTCAGGTAGAATCA
-----+-----+-----+-----+-----+-----+-----+ 1080
GTACTTTAAACCTTTATAGGTAAACAAACTCTTTTAGGAGTCGGACAAGTCGATCTTAGT
-----+-----+-----+-----+-----+-----+-----+

1081 CTGTCAAAACCATGTGGTTTTTGATAATCACTCGCCCATTTTAGCTAAGAAATTTGTGTAAAG
-----+-----+-----+-----+-----+-----+-----+ 1140
GACAGTTTGGTACACCAAAAACCTATTAGTGAGCGGTAAATAATCGATTCTTTAAACAACATTC
-----+-----+-----+-----+-----+-----+-----+

1141 AAACGCCCTGCGAAGCAGTTAGATGGAGTCTTTCCAGATATCTTTTGAATGAGGCGAG
-----+-----+-----+-----+-----+-----+-----+ 1200
TTTCCGGACGCCCTTCGTCAATCTACCTCAAGAAAGGCTATAGAAAAAACTTACTCCGCTC
-----+-----+-----+-----+-----+-----+-----+

1201 CAATTTGACCGCTGACTTGATACCGAGTTATTTCTGTCTACATCCAAATAGGCTTCGT
-----+-----+-----+-----+-----+-----+-----+ 1260
GTTAAAACTGGCGACTGAACATATGGCTCAAAATAAAGACAGTGTAGTTTATCCGAAGCA
-----+-----+-----+-----+-----+-----+-----+

1261 CAATGCTCATGGGTTCAATCAAAATCTGTATAGCGCTTAAATAATAGCTCGAATCCGGAGTC
-----+-----+-----+-----+-----+-----+-----+ 1320
GTTACGAGTACCCAAAGTTAGTTTAGACATATCCGGAATTTTATTCGAGCTTAGGCCCTCAG
-----+-----+-----+-----+-----+-----+-----+

1321 CCACAGACTTGTATTTCTCATATAATTCCTGAGATAAAGACAGCCCTGGGACACAGCTTCAT
-----+-----+-----+-----+-----+-----+-----+ 1380
GGTGTCTGAACATAAAGAGTATTAAGGACTCTATTCTGTGCGGACCCCTGTGTGCAAGTA
-----+-----+-----+-----+-----+-----+-----+

1381 AAGCTTCCTTGGAACTCATGGCAGAAATGGACACCAAAAGCTCTTTCCTCATTAACACTACAG
-----+-----+-----+-----+-----+-----+-----+ 1440
TTCCGAAGGAACCTTGAGTACCGTCTTACCTGTGGTTTTTCGAGAAGGAGTATGTGATGTC
-----+-----+-----+-----+-----+-----+-----+

PCT/US05/27239

254/487

Figure 101C

1441 TAGAACGACTCCCGCTCCACCTGTTTGGCGAGGTCGCTTCCAATAATGACAGGTTTTC
-----+-----+-----+-----+-----+-----+-----+
1500 ATCTTTGCTGAGGGCAGGTGGACAAACGGCTCCAGCGAAGGTTATTTACTGTCCAAAAG
-----+-----+-----+-----+-----+-----+-----+
1501 CTCTGAGTTTAGGATTATCCCTGATTCCACTGCAGCAAAAAGGCATCCATGCAATAT
-----+-----+-----+-----+-----+-----+-----+
1560 GAGACTCAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTTTTCCGTAGGTACAGTTATA
-----+-----+-----+-----+-----+-----+-----+
1561 GGATGATTTTCTCTTGACAAATCATTTAACAAAGGAAAAATCAACATGCCCTAGCACCTTTT
-----+-----+-----+-----+-----+-----+-----+
1620 CCTACTAAAAGAACTGTTTAGTAAATTGTTTCCCTTTTAGTTGTACGGATCGTGGAAAA
-----+-----+-----+-----+-----+-----+-----+
1621 TATACTCTTCGAAAAATCTCTTCAAAACCGTCAGCTTCCATCTGCAACCTCAAAACAGTA
-----+-----+-----+-----+-----+-----+-----+
1680 ATATGAGAAGCTTTTAGAGAAGTTTGGTCAGTCGAGGTAGACCTTGGAGTTTGTGCTAT
-----+-----+-----+-----+-----+-----+-----+
1681 TTTTGAGCTGACTCGTCAGTTCTATTATACAACTCAAGCAGTCGTTTGACGAGCCTGC
-----+-----+-----+-----+-----+-----+-----+
1740 AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTACAGAACTCGTCGGAGG
-----+-----+-----+-----+-----+-----+-----+
1741 GGCTAGTTTCTAGTTTGCTTTTTCGATTTCCTCATTTGAGTGTAACCTGCTTATTTTCTTTTAT
-----+-----+-----+-----+-----+-----+-----+
1800 CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACACATTCAGCAATTAAGAAATA
-----+-----+-----+-----+-----+-----+-----+
1801 TATACCTTTTTC TGAAAAAAGAAAAAGGACTTTATTTTTCATAAATAATATATACA
-----+-----+-----+-----+-----+-----+-----+
1860 ATATGGGAAAAAAGACTTTTTCCTTTTTCCTGAAATAAAAAAGTTTATATATATGT
-----+-----+-----+-----+-----+-----+-----+
1861 GTTTGAAATAAATAATAGACTGTTTGTAGAAAAAGTGTAAAAATAGGAATTTTTCACCT
-----+-----+-----+-----+-----+-----+-----+
1920 CAAACTTTATTTTATATCTGACAAATCTTTTCTTTTTCACATTTTATTCCTTAAAAAGTGA
-----+-----+-----+-----+-----+-----+-----+

Figure 101D

1921 TGTGAAATCGGTACTTTATGGTATACTTGTCTCANGAATGTAACAGATGACTGTTACT + 1980
-----+-----+-----+-----+-----+-----+-----+-----+
ACAACTTTAGCCAAATGAATACCATATGAACAGAGTACTTACATTGTCTACTGACAAATGA
-----+-----+-----+-----+-----+-----+-----+-----+
AGAAAAAGAGGACATTAATATGTTGTTAAGACAGTTGTTGAAGCACAAAGATATTTTTC + 2040
-----+-----+-----+-----+-----+-----+-----+-----+
TCCTTTTTCCTCTGTAATTATACCAACAATCTGTCAACAACACTGCTGTTCTATATAAAAC
-----+-----+-----+-----+-----+-----+-----+-----+
c M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate
acetyltransferase (pfl). It is out of the
pilus locus
2041 ACAAGCTTGGGAAGGCTTCAAGCGTAGATGGAAGAAAAAGCAAGTATCACGCT + 2100
-----+-----+-----+-----+-----+-----+-----+-----+
TGTTTCGAACCCCTTCCGAAGTTTCCGCATCTAACCTTCTTTTTCGTTTCACATAGTGCGA
-----+-----+-----+-----+-----+-----+-----+-----+
c K A W E G F K G V D W K E K A S V S R F -
2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCTTCAGGACCAACAG + 2160
-----+-----+-----+-----+-----+-----+-----+-----+
AACATGTTCCGATGATGTGTGGAATACTACCTCTGCTTTCGAGGAACGCTCGTTGTC
-----+-----+-----+-----+-----+-----+-----+-----+
c V Q A N Y T P Y D G D E S F L A G P T E -
2161 ACGGTTCACTTCACATCAGAAAAATTGTAGAAGAACTAAAGCACACTAGCAAGAAATC + 2220
-----+-----+-----+-----+-----+-----+-----+-----+
TCGCAAGTGAAGTGTAGTCTTTTAAACATCTTCTTTGATTTCTGCTGTGATGCTTCTTTGAG
-----+-----+-----+-----+-----+-----+-----+-----+
c R S L H I K K I V E E T K A H Y E E T R -
2221 GTTTCCTCAATGGACACTGCTCCAACATCTATCGCTGATATCCCTGCTGGATTATCGACA + 2280
-----+-----+-----+-----+-----+-----+-----+-----+
CAAAGGTTACCTGTGACAGGTTGTAGATAGCGACTATAGGACGACCTAAATAGCTGT
-----+-----+-----+-----+-----+-----+-----+-----+
c F P M D T R P T S I A D I P A G F I D K -
2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTTCAAAATGGAACCTTCATGC + 2340
-----+-----+-----+-----+-----+-----+-----+-----+
TTCCTTTACTTCAATAGAAACCATAGGTTTACTACTTGAAGATTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -
2341 CAAAAGGTGGTATCCGTATGGCTGAACTACTTTAAAGAAATGGATACGAACCGAGACC
-----+-----+ 2400
GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTCTTTTACCTATGCTTGGTCTGG
-----+-----+
C K G G I R M A E T T L K E N G Y E P D P -
2401 CAGCTGTTACGAAATCTTCACTAAATATGTAACAACAGTTAACGACGGTATTTCCGCTG
-----+-----+ 2450
GTCGACAAGTGTCTTTAGAAAGTGATTTATACATTTGTTGTCAATTGCTGCCATTAAGGCAC
-----+-----+
C A V H E I F T K Y V T T V N D G I F R A -
2461 CCTACACTTCAAAATATTCGTGCGGCTCGTCAAGCACACACTGTAACCTGGTCTTCCAGATG
-----+-----+ 2520
GGATGTGAAGTTTATAACGACGCGGACGACGTACGTTGTGTGACATTGACCAGAAAGGCTTAC
-----+-----+
C Y T S N I R R A R H A H T V T G L P D A -
2521 CATACTACCGGACGATATCATCGGTGTTTACGACGCTTCTCTCTTTACGGTGCAGACT
-----+-----+ 2580
GTATGAGTGGCCTGCATAGTAGCCACAATGCGTGCAGAACGAGAAATGCCACGCTCTGA
-----+-----+
C Y S R G R I I G V Y A R L A L Y G A D Y -
2581 ACTTGATGCAAGAAAAGTAACGACTGGAATGCAATCAAGAAATCGATGAAGAAACAA
-----+-----+ 2640
TGAACACTGCTCTTTTTCATTGCTGACCTTACGTTAGTTCTTTTAGCTACTTCTTTGTT
-----+-----+
C L M Q E K V N D W N A I K E I D E E T I -
2641 TCCGTCTTCGTGAAGAAGTAAACCTTCAATACCAAGCATTCGAACAAGTTGTTCCGCTGG
-----+-----+ 2700
AGGCAGAACCACTTCTTTCATTGGAAGTTATGTTCTGTAACGTTGTTCAACAAGCGGACC
-----+-----+
C R L R E E V N L Q Y Q A L Q Q V V R L G -
2701 GTGACCTTTACGGGTTGATGTTCCGCAACCCAGGATGAACGTGAAGAGCAATCCAAAT
-----+-----+ 2760

257/487

Figure 101F

CACTGGAAATGCCCAACTACAAGCGTTTGGTCGCTACTTGCACCTTCTTCCTTAGGTTA
D L Y G V D V R K P A M N V K E A I Q W -
GGGTTAACATTGCTTTTCATGGCTGTCTGCGGTGTGATTAACGGTGTCTACATCTCTAG + 2820
-----+-----+-----+-----+-----+-----+
CCCAATTGTAACGAAAGTACCGACAGACGGCACACACTAATGCCACACGATGTAGAGATC
V N I A F M A V C R V I N G A A T S L G -
GTGCTGTACCAATCGTATTGGACATCTTTGACAGAACGTGACCTTCTCGTGGTACATTTA + 2880
-----+-----+-----+-----+-----+-----+
CAGCACATGGTAGCATTAACCTGTAGAAAGCTCTTGGCTGGAACGAGCACCATGTAAAT
R V P I V L D I F A E R D L A R G T F T -
CTGAATCAGAAATCCAAGAAATTCGTTGATGATTTCGTTATGAAACTTCGTACAGTTAAAT + 2940
-----+-----+-----+-----+-----+-----+
GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATCTCAATTTA
E S E I Q E F V D D F V M K L R T V K F -
TTGCTCGTACCAAGCTTATGACCAATTGTACTACTAGGTGACCCACCTTTATCACAACCTT + 3000
-----+-----+-----+-----+-----+-----+
AAGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACTGGGTGGAAATAGTGTGAA
A R T K A Y D Q L Y S G D P T F I T T S -
CTATGGCTGGTATGGGTAAACGACGGTCTGTCACCGTGTACTAAGATGGACTACCGTTTCT + 3060
-----+-----+-----+-----+-----+-----+
GATACCGACCATACCCATTTGCTGCCAGCTGCGCACATGATCTCTACCTGATGGCAAGA
M A G M G N D G R H R V T K M D Y R F L -
TGACACTCTTGACAAACATCGGTAACTCACCAGAACCAAACTTACAGTTCTTTGGACTG + 3120
-----+-----+-----+-----+-----+-----+
ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAAGTGTCAAGAAACCTGAC
N T L D N I G N S P E P N L T V L W T D -

PCT/US05/27239 258/487

Figure 101G

3121 ~ACAAATGCCATACAACTCCGTCGCTACTGTATGCACATGAGCCACAAACACTCTCTTA
+-----+-----+-----+-----+-----+-----+ 3180
TGTTTAAGCGTATGTTGAAGGACGCGATGACATACGTGTACTCGGTGTTGTGAGAAGAT
+-----+-----+-----+-----+-----+-----+
K L P Y N F R R Y C M H M S H K H S S I -
+-----+-----+-----+-----+-----+-----+
TCCAAACGAAGGTGTAAACAACATGGCTAAAGACGGATATGGTGAATGAGCTGTATCT
+-----+-----+-----+-----+-----+-----+ 3240
AGGTTATGCTTCCACATGTTGTTACCGATTCTCTGCCCTATACCACCTTACTCGACATAGA
+-----+-----+-----+-----+-----+-----+
Q Y E G V T T M A K D G Y G E M S C I S -
+-----+-----+-----+-----+-----+-----+
CATGCTGTGTCTCCACTTGATCCAGAAAATGAAGAACACGCCACACATCCAGTACT
+-----+-----+-----+-----+-----+-----+ 3300
GTACGACACACAGAGGTGAACGTAGGTCTTTTACTTCTTGTGCGGTGTGTAGTTCATGA
+-----+-----+-----+-----+-----+-----+
C C V S P L D P E N E Q R H N I Q Y F -
+-----+-----+-----+-----+-----+-----+
TCGGTCTCGTGTAAACGTTCTTAAAGCCCTTCTTACTGTTTGAATGGTGGTTACGACG
+-----+-----+-----+-----+-----+-----+ 3360
AGCCACGACACATTTGCAAGATTTCGGGAAGAAATGACCAAACTTACCACCAATGCTGC
+-----+-----+-----+-----+-----+-----+
G A R V N V L K A L L T G L N G G Y D D -
+-----+-----+-----+-----+-----+-----+
ATGTTACAAAGACTACAAAGTATTTGATATCGAACCAATCCGTGACGAAGTTCTTGAAT
+-----+-----+-----+-----+-----+-----+ 3420
TACAGTGTCTCTGATGTTTCTATAAACTATAGCTTGGTTAGGCACCTGCTTCAAGAACTTA
+-----+-----+-----+-----+-----+-----+
V H K D Y K V F D I E P I R D E V L E F -
+-----+-----+-----+-----+-----+-----+
TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTTGTACTGGTTGACTGACACTTACGTAG
+-----+-----+-----+-----+-----+-----+ 3480
AACTTAGTCAATTCGCTTGAACCTTTTATAGAGAACTGACCAACTGACTGTGAATGCATC
+-----+-----+-----+-----+-----+-----+
E S V K A N F E K S L D W L T D T Y V D -
+-----+-----+-----+-----+-----+-----+
ATGCTTGAACATCATCCACTACATGACTGATAGGTACAACATCGAAGCTGTTCAAATGG
+-----+-----+-----+-----+-----+-----+ 3540
TACGGAACCTTGTAGTAGGTGATGTACTGACTATCCATGTTGATGCTTCGACAAGTTTACC
+-----+-----+-----+-----+-----+-----+

PCT/US05/27239 259/487

Figure 101H

C A L N I I H Y M T D R Y N Y E A V Q M A -
3541 CCTTCTGCGCAACTAAACAACGTCGCCAACATGGGATTCGGTATCTGTGGATTGCTTAACA
+-----+ 3600
GGAAGAACGGTTGATTGTTGACCGTTGTACCCCTAAGCCATAGACACCTAAACGATTGT
+-----+
C F L P T K Q R A N M G F G I C G F A N T -
3601 CTGTTGATACATTGTCAGCTATCAAAATACGCTACAGTTAAACCAATCCGTCGACGAAGATG
+-----+ 3660
GACAACTATGTAACAGTCGATAGTTTATGCGATGTCAATTTGGTTAGGCACCTGCTTCTAC
+-----+
C V D T L S A I K Y A T V K P I R D E D G -
3661 GCTACATCTACGATTACGAACAATCGGTGACTACCCACGCTGGGGTGAAGATGACCCAC
+-----+ 3720
CGATGATGATGCTAATGCTTTGTTAGCCACTGATGGTGGCGACCCACTTCTACTGGGTG
+-----+
C Y I Y D Y E T I G D Y P R W G E D D P R -
3721 GTTCAACGAATTGGCAGAAATGGTTGATCGAAGCTTACACAACCTCTCTACGTAGCCACA
+-----+ 3780
CAAGTTGCTTAACCGTCTTACCACTAGCTTCGAATGTGTTGAGCAGATGCATCGGTGT
+-----+
C S N E L A E W L I E A Y T T R L R S H K -
3781 AACATACAAAGACCGCAGAGCTACAGTATCCTTTTACAAATCAGTCTACGTTGCTT
+-----+ 3840
TTGATATGTTTCTGCGTCTTCGATGTCATAGTGAAACTGTTAGTTAGATTCCAAACGAA
+-----+
C L Y K D A E A T V S L L T I T S N V A Y -
3841 ACTCTAAACAACTGGTAACACCAAGTTCACAAAAGTGTATACCTCAACGAAGATGGTT
+-----+ 3900
TGAGATTGTTTACCATTTGAGTGGTCAAGTGTTCACATATGGAGTTGCTTCTTACCAA
+-----+
C S K Q T G N S P V H K G V Y L N E D G S -
3901 CTGTGAACCTTGCTAACTTGAATCTTCTCACCAGTGTCTAACCCATCTTAAACAAAGCTA
+-----+ 3960

PCT/US05/27239 260/487

Figure 101I

GACACTTGAACAGATTGAACTTAAGAAGAGTGTGTCACGATGGGTAGATTGTTTCGAT
V N L S K L E F S P G A N P S N K A K -
AAGTGGTGGTTGCAAAACTTGAACCTCAGTCTTCTAGCCCTGACTTTAGTTATGCACGTG + 4020
TTCCACCAACCAACGTTTGAACCTTGAAGTGAAGATCGGAACGAAATCAATACGTCGAC
G G W L Q N L N S L S L D F S Y A A D -
ACGGTATCTCATTTGACTACACAAGTATCACCTCGCGCTCTTGGTAAGACTCGTGTGATGAAC + 4080
TGCCATAGAGTAACCTGATGTGTTTCATAGTGGAGCGCGAGAACCATTCCTGAGCAGCTACTTG
G I S L T T Q V S P R A L G K T R D E Q -
AAGTGTATACTTGGTAACAATCCTTGATGGTTACTTCGAAAACGGTGGACAAACAGCTTA + 4140
TTCAACTANTGAACCATGTTAGGAACCTACCAATGAAGCTTTTGCCACCTGTTGTGCAAT
V D N L V T I L D G Y F E N G G Q H V N -
ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTCAGGCGAAGACGTTA + 4200
TGAACCTGCAATACCTGAACCTTGCTACAAAATGCTTTTATAGTACAGTCGCGCTTCTGCAAT
L N V M D L N D V Y E K I M S G E D V I -
TCGTACGTATCTCTGGATACCTGTATAACACATAAATACCTCAGTCCAGAACAAAAACTG + 4260
AGCATGCATAGAGACCTATGACACATTTGTGATTTATGGAGTGAAGTCTGTTTGTGAC
V R I S G Y C V N T K Y L T P E Q K T E -
AATTGACACAACGCTGCTTCCAGGAAGTTCCTTCAATGGATGACGCCCTTGGATGCAATTGA + 4320
TTAACTGTGTTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAAC
L T Q R V F H E V L S M D D A L D A L S -

261/487

Figure 101J

4321 GCTAATCAAGTCTTGAATAATAAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA 4380
-----+-----+-----+-----+-----+-----+
CGATTAGTTCAGAAGCTTATTTATTTTCCCGAGAAACAGATTGACATCACCCAACTCTCTTT
c *
4381 AGCTAAGCTCGAGAAAGGACAAATTTGTCTCTTTTGTATGTTTCAGAGCGATGAAA 4440
-----+-----+-----+-----+-----+-----+
TCGATTCGAGCTCTTTCCTGTTTAAACAGGAAAGAAAAAACTACAAGTCTCGCTACTTT
e * A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase
4441 ATCCGTTTTTTGAAGTTTCAAAAGTTCCGAAACCAAGGCAATTCGCTTGATGCTTTTG 4500
-----+-----+-----+-----+-----+-----+
TAGGCAAAAAAAGTTCAAAAGTTTCAAGGCTTTTGGTTTCCGTTACCGGAACACAGAAAAC
e I R K K F N E F N R F G F A N R K I D K -
4501 ATGAGTTTGTAGTGGCTCAAGTTTAGCGTTAGATAAAGGCAATTCATGGCGTTAGTG 4560
-----+-----+-----+-----+-----+-----+
TACTCAACAAATCACCGGAGTTCAATCGCAATCTTATTCCTGTTAAGTTACCGCAATCAC
e I L K N T A E L K A N S Y P L E I A N T -
4561 ATGTAGTTTTTATAGCAATAAATGTGCTCAAAGTGGTTTAAAGGTGCGTTGAGATGA 4620
-----+-----+-----+-----+-----+-----+
TACATCAAAATATCGTTTATTTACACGAGTTTCACCAAAATTTCCACGCCAACTCTACT
e I Y N K Y C I F T S L T T K F T R N L H -
4621 GGTAACGTGCTTGAATTAAGCCCCAAACTGGTCAGTATTCCTCTTGTAGATGAAAT 4680
-----+-----+-----+-----+-----+-----+
CCATTGCACAGAACTTAATTCGGGGTTTGACCAATCATAGAGAGAAATCTACTTTTA
e P L T D Q I L G W F Q D T N K E Q L H F -
4681 AGGAGTAGTTGATACAGGTCATAGTAATCTTTAAGTTCAGGTACTAGAGTAAAGATTTTC 4740
-----+-----+-----+-----+-----+-----+
TCCTCATCAACTATGTCAGTATCATTAGAAAATTCAGTCCATGATCTCATTTCTAAAG

262/487

Figure 101K

e L L L Q Y L D Y Y D K L E P V L T F I K -
4741 TTGAGACACTCCCTAGGAGTTAAGGTCCTCTGTGAAAGTTCTAGCATAGAAAGGCTTAAGA + 4800
AAGTCGTGAGGGATCCCTCAATTCAGAGACATTCAGATCGTATCTTTCGGAATTC
K L C E R P T L T E R F T R A Y F P K L -
4801 GAGAGTTCCGACTATCTTTAGGATAAATTTCCAGTAATATTTAAGAGCCTGTGATTCC + 4860
CTCTCAAAGGCTGATAGAAAATCCCTATTAAAGGTCATTATAAATTCGAGACATAAGG
S L K R S D K L I F K W Y Y K L A R Y E -
4861 AGAGATTATCATCAAAATTCCTTCATGATGTTGATTCTAGTCTGATTAAGACCCCTGCTC + 4920
TCTCTAAATAGTAGTTAAAGGAAGTACTACAACCTAAGATCAGACATAATTCGCGGACGAG
L S K D D F Q K M I N I R T Q N L A R S -
4921 ATGTGTTGGACAATGTGGAACGATCGAGAACAAATTTAGCATGGGAAATAATTTCTTA + 4980
TACACAACCTGTTACACCTTGTAGTCTGTTGTTAAATCGTAACCTTTTATTAAGAAAT
M H Q V I H F R D L V I K A N P F L K K -
4981 ATGAGAGGATATAACTTCCAGACATATCAACAGTACGACTTAACTTTTCTTCTAGCT + 5040
TACTCTCCCTATATTGAAGGCTCTGTATAGTTGTCACCTGCTGAAATTGAAAAAAGATCGA
I L P I Y S G S M D V T V V K V K K R A -
5041 TCTTTCGAGTACTTGAAGAAATGATTTCGGATGGTTGTTTGGACGCTCTGTATCAAGAATG + 5100
AGAAAGCTCATGAACCTTCTTACTAAAGCCACCAACAACTGCAGACATAGTTCTTAC
E K S Y K F F H N R I T T Q R R N D L I -
5101 GTCATGATTTTCTTAGTGTGAAATCCCTGAGCAATGAAGCAATTTCCCTTCTGCTAG + 5160

Figure 101L

e CAGTACTAAAGAAATCACAACCTTTAGGACTCGTTACTTTTCGGTTAAAGGGAAGACCATC
T M I K K T N F D Q A I F A L K G K Q Y -
5161 GAGAAATTCATCCCGAGGAGAGATTTCAGGCAAGTGGTGAATCCTCTTGGAAATGAAT
CTCTTAAGTAGGTCCTCTCCTAAAGTCCGTTTCACCATAGGAGAACCTTTTACTTTA
S F E D W S L I E P L T T Y D E Q F H F -
5221 TGCTTGAGCTTACGATAGACGGTAGAGGTAGAGTAGGGCTAATTTAGAACGATA
ACGAACTCGAATGCTATCTGCCATCTCCATCTCCATCTACCGATTAAATCTTCGCTAT
Q K L K R Y V T S T S I A L K S A I -
5281 TGTGTAAGAGCCCTCTCTGTGAGTAGGAGTTGGGCAATTTCTGTCTCACCATTCCGAG
ACACATTCCTCGGAGAGACAACCTCATCTCAACCCGTTAAAGACAGAGTGGTAAAGGCTC
H T L A E R N L L L Q A I K Q R V M E S -
5341 ATTTGGCAATTTTCTGAAGAGAGTGTGTTTCAGCTACAGTGACTTCCGACAGGACTTG
TAAACCGTTAAAGACTTGCTCTCAACAAGTCGATGTCACCTGAAAGGCTGTCCTGAAC
I Q C N K Q V L T T E A V T V K R C S K -
5401 CATTGAAATCGTCTCTTTTCAAAATGAATGAGGCTAGGAAACCAATCTCGATAAAA
GTAACCTTAGCAGAGAAAAAGTTACTTACTCCGATCCCTTTGGTGGTTAGAGCTATTTT
C Q F R R K K L H I L S P F G G I E I F -
5461 GGGATTTTAGAGGCTTTTGGAGTCGTATTGATTTGTTTTCCTTTACAGTGTTTACAT
CCCTAAATCTTCGGAAAAACCTTCAGCATAAACTAAACAAAAGGAATGTCACAAAATGTA
P I K S P K Q F D Y K I Q K G K C H K C -
5520

Figure 101M

TTAGTGGTGATAATCAAGTGTAGCGAAGACTTCGATATGGGTATCGTCTGAATGGCT 5521
-----+-----+-----+-----+-----+-----+ 5580
AATCCACCCACTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCAGGACTTACCGA
K P P H Y D L T A F V E I H T D H Q I A -
TTATTTAAGGTGATGTTTTTGTCTTTTATTCGATGAGTAATGTGGTATGATGTGT 5581
-----+-----+-----+-----+-----+-----+ 5640
AATAAATCCCACTACAAAAACAGAAAAATAAGGCTACTCATTACACCACTAATACTACACA
K N L T I N K D K I G I L L T T H N I H -
TCCATAAGATACTTTTCTAATGAGTTGTTTAGGGCTTTTTCATTATAAGTCTTATGGGACT 5641
-----+-----+-----+-----+-----+-----+ 5700
AGGTATTCATGAAAGATTACTCAACAAATCCCGGAAAAGTAATATTCAGAAATACCCCTGA
E M
TTTTTGATACTCAAAAAGCCCTTAATAATCTCCACAGTGGGATTTACCCACTACAGAAATTA 5701
-----+-----+-----+-----+-----+-----+ 5760
AAAACTATGAGTTTTTCGGGATATTAGAGGTGTCAACCCATAAATGGGTCATGCTTTAAT
TAGAGCCAGAAAAACACTTTTGTCTACTAGCAGAAACTAGAGAGCAGAGTGTTTTCT 5761
-----+-----+-----+-----+-----+-----+ 5820
ATCTCGGTCCTTTTGTGAAAAACAAGTGTATCGTCTTTGATCTCTCGTCTTCACAAAAAGA
GTTCAGATTTACCCAAAACTGGGAAATATGGGGATAAGAATAGAGATGGCTTAGGAAGCC 5821
-----+-----+-----+-----+-----+-----+ 5880
CAAGTCTAAATGGGTTTGTACCCCTTATACCCCTATTCCTTATCTCTACCGAATCCTTCGG
CCTTTTGTGTGTAGACAGTACGATGAACCTTATAACAAATAGTGAGCCCTTTTACGAATC 5881
-----+-----+-----+-----+-----+-----+ 5940
GGAAAAACACAGATCTGTCTACTTGTGAATATTGTTTATCTACTCGGAAAAAATCGTTAG
* L L Y H A K K A I -orf2_670 homologue of sp0461, transcriptional
regulator
ATTGCCACCCGTTTGTCAAAAGCCCTCTTTTTCGGATATCTACAATGTCTGATAGATGAGA

$$265/487$$
[illegible]

PCT/US05/27239

266/487

Figure 1010

```
6361 TCCTGAACAATTGCTTTTCGAAATATGATACAGTGGCTTGTGCTTTCATCCCATTAATG+ 6420
-----+-----+-----+-----+-----+-----+-----+
AGAACTTGTTAACGAAAGCTTTATATACTATGTCACCGACAGCGAAGTTAGGGTATTACA
e E Q V I A K S I H Y L P K D S E I G Y H -
6421 TCGTAATAATTATAATAGGGAAGTATTTGTAACCAAAACAAAACGTTCTTGTTAAG+ 6480
-----+-----+-----+-----+-----+-----+-----+
AGCAATTAATAATATATCCCTTGATCTAAACATTTGGTTTGTGTTTTCGAAGAACAAATTC
e E Y Y N Y Y P V L N Q L G F L F T R T L -
6481 AAAGTCAGTGTCTTAAAGAAAGAGAAATTCGAATGTCATTTCCCTAAGATATCTTG+ 6540
-----+-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAATTTTCTCTTCTTAAGCTTTACAGTAAGGATTCCTATAAGAAC
e F T L A T L F S L S N S I D N G L I N K -
6541 AACTGGATAGTAGATGCTTTCCTCTTGTATGCTGAAGAATCAGTGAATAGTATGATC+ 6600
-----+-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCACTTATCATCTACTCAG
e F K S L L H K G R T H Q L I L Q I T H T -
6601 TTTTTCCTTGATTCATTTTGTCTTGGAAAACGAAGAATTAGCAGAACATAAACCAA+ 6660
-----+-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAAACAGGAACCTTTGCTTCTTAATCGTCTTGTATTGTTT
e K K E Q N W K D K S F S S N A S C Y V L -
6661 AAGATATAATCCAGTTCCTCTGAGTAAAGTCATGTTGGCATGTTGGCTCTAAGTAAGTT+ 6720
-----+-----+-----+-----+-----+-----+-----+
TTCATATATTAGTCAAGAAGGACTCATTTTCAGTACACCGTACACCGAGATTCATCAA
e F I Y D L E E Q T F T M N A H P E L Y T -
6721 TGGCAATGTTCCATCAAAATCGGATACATAAAGAGTTTAAATTTTCAAACTCTTTG+ 6780
-----+-----+-----+-----+-----+-----+-----+
ACCGTTACAAGTAGTTTACCTATGTTATTTCTCCAAAAAATTAAGTTTGAGAAAC
```

Figure 101P

```

e      Q C H E M L I P Y M F L N K L K E F E K -
6781  GACTCAGGAACTCAAGTGGAAATTCCCGACGCTTTCCAAGTGAGTCCACATGATGCTA
      +-----+-----+-----+-----+-----+
      6840
      CTGAGTCCCTGTGAGTTCACCTTTAAGGCGCTGAAAGGTTCACTCAGCGTGATCATACGAT
      +-----+-----+-----+-----+-----+
e      S E P F E L P F E R R K W T L A V L I S -
      6841
      AAATGAACATACTCGTTCAGGTGTGATTTCTAACAGTTCACTGACTGAGTTGACAATTAGAC
      +-----+-----+-----+-----+-----+
      6900
      TTTACTTGTATGAGCAGTCCACACATAAGATGTCAAGTACTGACTGACTCAACCTTAAATCG
      +-----+-----+-----+-----+-----+
e      F H V Y E D P T I E L L E H S L Q S N S -
      6901
      TGCACAATCATATGTGTGACCCCAATCCATACTTCCATCATTCAAATCATATAATCTCAATA
      +-----+-----+-----+-----+-----+
      6960
      ACGTGTTAGTATACACACTGGTTPAGGTATGAAGGTAGTAAAGTTAGTTAGTTAGAGTTAT
      +-----+-----+-----+-----+-----+
e      Q V I M H T V W D M S G D N L D Y I E I -
      6961
      CCAAAATGAAACTGGAGGAGTGCAATTAAAAACGAATCGGATATTCAGGACCAACTACT
      +-----+-----+-----+-----+-----+
      7020
      GGTTTTACTTTGACCTCCTCAGGTTAATTTTTTGTCTTACGCTATAAGTCCGTGTTGATGA
      +-----+-----+-----+-----+-----+
e      G F H F Q L L A I L F R I R Y E P G V V -
      7021
      TGATTTTTCACAAGGTCCAAACCTTACTGAAGCTAGTAACAAGCCACACTTTTGTCTGACG
      +-----+-----+-----+-----+-----+
      7080
      ACTAAAAAGTGTCCAGGTTTGGATGACTTGCATCATTTGTTGGGTGTAAGAACAGCATGC
      +-----+-----+-----+-----+-----+
e      Q N K V L D L G V S R L L L G C K Q R V -
      7081
      CGGTAGCCTGTTGCGATGGAATAATACTCTTTTGTGTGTAATAATCGTTAAAGCTTTGATTA
      +-----+-----+-----+-----+-----+
      7140
      GCCATCGGACAACGCTACCTTTATATGAGAAAAACACATTTAAGCAATTTGCAACATAAT
      +-----+-----+-----+-----+-----+
e      R Y G T A I S I Y E K Q T F E N F S Q N -
      7141
      CCTTGTTAGTAGAAAGCGGACTATTTTAAAAATAGTTGATTGGTTATATAAGCTGATGG
      +-----+-----+-----+-----+-----+

```

268/487

Figure 101Q

7201
e G Q L L F F R L I K L I T S Q N Y L Q H -
GGAACATCATCTTCTTGGCCATATAAAATTTTATCAACTAACCAATATTTGACTACC
AAGTAATAATTCTGTTGATGAGAAATGGTTCGATTAAATTGAACCTGTTGCGTATCTAAA
TTCTATTATTAAGCAAACTACTCTTACCACAGCTAATTAACCTTGAAACAACGCATAGATTT
F Y Y N T Q H S H E I L Q V Q Q T D L -
TTAAATGTCACCTCTCCGAAATGTTCTTGTAATTCCTGCAAAATGCTTAGGAGACTT
AATTTACAGTTGAGAAGGAGCTTACAAGAACATTAAGGACGTTTACGAATCCTCTGAA
N F T L E E E F T E Q L E Q L I S L S -
TTAGATTGTAATGAAGTTAAAGTAGACAGTTCATCTAGTCAATAGACCGAATATCCAAT
AATCTAACATTACTTCAATTTTCATCTGTCAAGTAGATCAAGTTATCTGGCTTATAGGTTA
K S Q L S T L T S L E D L E I S R I D L -
AATATATTATAAAGGTAATTTTATCTGTGAATCTTTTTCATCTGATTTGTTTAGCATA
TTATATATAAATTTTACCATTAAATATAGACATTAAGAAAAGTTACATAAACAAATCGTAT
L I N L I T I K D T I R K E I Y K N L M -
GTACCGAATCTTAGTTGCATATAGATAATTTTAATTATTAATACAAAAGAACTAAT
CAATGGCTTAGAATCAACGTTATCTATTAAAAATTAATAATATTGTTTTCTTTGATTA
TGCTCTGTCAAAAAGGTTGTGGAATTTCCGACTTTTATTGATAAAACAGCATGTAATAAAA
ACAGAACAGTTTTCCTCCACACCTTAAAGGCTGAAATACTATTTTGTCTGATATTATTTT
GGCATTTTAAAGATAGTAATGAGTATTTGGTGGAGTTTATGCTTATTTTATTTATAGA
7260
7320
7380
7440
7500
7560
7620

CGGTAAATAATTTCTATCATTAATCCATACCCACCTCAAAAATACCGAATAAAAAAATAATCTT
7621
AAATATTTTTTTTATCAAAATATTTGTCGTTCTATATAAAAAATATGTGATAAAAATATCTATT
TTTATAAAAAAATAGTTTATAACACGCAAGATATTTTTTTTATACACTAATTTTTTATAGATAA
7680
GTGATGGAAGTTGTTTTTAAATTTATACAGGATAGTTAATAGTAATACTATACATACTAT
7681
CCTACCTTCAACAATAATTAATAATGATCCTATCAATTAATGATATGATATGATATGATA
7740
ATGTATACAAAGTGTGTCATGTCACGTTGAGAAAGATAGCTATTAACGCACCTTTTATACGC
7741
TAACATATGTTTCACACAGTAACGGTCCAACTCTTCTATCGGATATGCGTGAAATAATGCG
7800
TTTTGCTACGTTGTTAGTGAACGGATTAACCTCAGTGAGATAAATTTTATCAGAACATAA
7801
AAACGATGCAAAACAATCAGTTGCCTAATTGAGTCACCTCTATTTTAAATAGTCTTGTATT
7860
GTAATCCGTTTCTTCGTGTATACAGATTGAAAGTACCTATGAATCATAGAAGGATTAAC
7861
CATTAGGCAAGAAGACATATGTCTAACCTTTCATGGATACCTAGTATCTTCTCCAAATTGA
7920
TGTTCTATGAATAATGCTTAACAGGGAGACACACATGAAAAAGTAAGAAAGATATTCA
7921
ACAAGATCTTATTAAGAAATGTCCTCTGTGTACTCTTTTTCATCTTCTTCTATAAAGT
7980
GAAGGCAGTTTCAGGACCTGCTGTATATCTCAGTTGACAGCTTTTTTCTCGATAGTTGC
7981
CTTCCGCTAACGTCCTGACACGACATATAGAGTCAACTGTGAAAAAAGAGCTATCAACG
8040
K A V A G L C C I S Q L T A F S S I V A -

270/487

8041	b	TTTAGCAGAAAACGCCTGAAACCGAGTCCAGCGATAGGAAAAGTAGTAGTAAAGGAGACAGG +-----+-----+-----+-----+-----+-----+-----+-----+ AAATCGTCTTTTGGCGGACCTTTGGTCAGGTCGGCTATCCCTTTTCATCACTAATATCCCTCTGTCC +-----+-----+-----+-----+-----+-----+-----+-----+ L A E T P E T S P A I G K V V I K E T G -
8101	b	CGAAGGAGGAGCGCTTCTAGGAGATGCCGTCTTGAGTTGAGTTGAAAAACAATACGGATGGCAC +-----+-----+-----+-----+-----+-----+-----+-----+ GCTTCCTCCTCGGAAGATCCTCTACGGCAGAAACTCAACTTTTGTGTTATGCCTACCGTGG +-----+-----+-----+-----+-----+-----+-----+-----+ E G G A L L G D A V F E L K N N T D G T -
8161	b	AACTGTTTCGCAAGGACAGAGCGCAAAACAGGAGAGCGGATATTTTCAAAACATAAAACC +-----+-----+-----+-----+-----+-----+-----+-----+ TTGACAAAAGCGTTTCCTGTCTCCGCGTTGTCTCTTCGCTATATAAAGTTTGTATTTTGG +-----+-----+-----+-----+-----+-----+-----+-----+ T V S Q R T E A Q T G E A I F S N I K P -
8221	b	TGGGACATACACCTTGACAGAAAGCCCAACCTCCAGTTGGTTATAAAACCCCTCTACTAAACA +-----+-----+-----+-----+-----+-----+-----+-----+ ACCCGTGATGTGGAAGTGTCTTCGGGTGGAGGTCAACCAATATTTGGGAGATGATTTGT +-----+-----+-----+-----+-----+-----+-----+-----+ G T Y T L T E A Q P P V G Y K P S T K Q -
8281	b	ATGGACTGTGTAAGTTGAGAAGAAATGGTCGGACGACTGTCCAAAGTGAACAGGTAGAAAA +-----+-----+-----+-----+-----+-----+-----+-----+ TACCTGACAACTTCAACTCTTCTTACCAGCCCTGCTGACAGGTTCCACTTGTCATCTTTT +-----+-----+-----+-----+-----+-----+-----+-----+ W T V E V E K N G R T T V Q G E Q V E N -
8341	b	TCGAGAAGAGGCTCTATCTGACCAGTATCCACAAACAGGAGCTTATCCAGATGTTCAAAC +-----+-----+-----+-----+-----+-----+-----+-----+ AGCTCTTCTCCGAGATAGACTGGTCATAGGTGTTTGTCCCTGAATAGGTCTACAAAGTTTG +-----+-----+-----+-----+-----+-----+-----+-----+ R E E A L S D Q Y P Q T G T Y P D V Q T -
8401	b	ACCTTATCAGATATTAAGGTAGATGGTTTCGGAAAAAAAACGGACAGACACAAGCGTTGAA +-----+-----+-----+-----+-----+-----+-----+-----+ TGGAAATAGCTAATAATCCATCTACCAAGCCTTTTTTTTCCCTGCTGCTGTTCCGCAACTT +-----+-----+-----+-----+-----+-----+-----+-----+ +-----+-----+-----+-----+-----+-----+-----+-----+

PCT/US2005/027239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -
TCCGAATCCATATGAACGTGTGATTCCAGAAGGTACACTTTCAAAGAGAATTTATCAAGT 8520
8461 +-----+
AGGCTTAGGTATACCTGCACACTAAGGTCTTCCATGTGAAAGTTTCTCTTAATAATAGTTCA
P N P Y E R V I P E G T L S K R I Y Q V -
GAATAATTGGATGATAACCAATATGGAATCGAGTTGACGGTTAGTGTAATAAACACGCGGT 8580
8521 +-----+
CTTATTAAACCTACTATTGGTTATACCTTAGCTCAACTGCCAATCACCATTTTGGTGTGCCA
N N L D D N Q Y G I E L T V S G K T T V -
TGAACGAAAGAAGCCTCTACTCCGCTAGATGTGTGTTATTCTATTAGATAACTCCAATAG 8640
8581 +-----+
ACTTTGCTTTCTTCGGAGATGAGCGGATCTACAACAATAAGATAATCTATTGAGGTTATC
E T K E A S T P L D V V I L L D N S N S -
TATGAGTAATATTCGACATAATCATGCCCATCGAGCGGAAAAAGCGGAGAAAGCGACACG 8700
8641 +-----+
ATACTCATATATAAGCTGTATTAGTAGGGGTAGCTCGCCTTTTTCGCCCTCTTCCGCTGTGC
M S N I R H N H A H R A E K A G E A T R -
AGCCCTTGTAGATAAGATTACCTCCAATCCAGATAATCGAGTAGCAGTGTGACTTATGG 8760
8701 +-----+
TCGGGAACATCTATTCTAATGGAGGTTAGGTCTATTAGCTCATCGTGAACACACTGAATACC
A L V D K I T S N P D N R V A L V T Y G -
CTCAACTATCTTTGACGGTTCAGAAGCTACTGTGGAAAAAGGGGTAGCAGATGCGAACGG 8820
8761 +-----+
GAGTGTAGAGAAACCTGCCAAGTCTTCGATGACACCTTTTCCCCCATCGTCTACGGCTGCCC
S T I F D G S E A T V E K G V A D A N G -
AAAAATATTGAATGACTCAGCTTTATGGACGTTTCGATCGTAGACGTTTACAGCTAAAC 8880
8821 +-----+

Figure 101U

TTTTTATAACTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAAATGCTGATTTTG
K I L N D S A L W T F D R T T F T A K T -
TTATAATTATAGCTTTTAAATCTCACATCAGATCCTACTGATATTCAAACTATTAAGGA 8940
8881 +-----+
AATATTAAATATCGAAAAATTTAGAGTGTAGTCTAGGATGACTATAAGTTTCATATAATTCCT
Y N Y S F L N L T S D P T D I Q T I K D -
TAGGATTCCATCAGATGCAGAGGAATTGAACAAAGACAAAATGATGATATCAATTCGGGCG
8941 +-----+ 9000
ATCCTAAGGTAGTCTACGTCCTCCTTAAGTTGTTTCTGTTTAACTACATAGTTAAGCCGGG
R I P S D A E E L N K D K L M Y Q F G A -
GACTTTTACCCAGAAAGGCTTTGATGACCGCTGATGATATCTTGACAAAGCAGGCAAGACC
9001 +-----+ 9060
CTGAAATGGGTCTTCCGAAACTACTGCGGACTACTATAGAACTGTTTCGTCCTGTTCTGG
T F T Q K A L M T A D D I L T K Q A R P -
AAACAGTAAAGGTTATTTCCACATTACAGATGGTGTTCGACTATGTCATATCCAAAT
9061 +-----+ 9120
TTTGTCATTTTCCAAATAAAGGTGAATGTCTACCACAAAGGCTGATACAGTATAGGTTA
N S K K V I F H I T D G V P T M S Y P I -
TAATTTTAAATATACAGGAACGACGCAATCGTACAGAACTCAGCTGAATATTTTAAAGC
9121 +-----+ 9180
ATTAAATTTATATATGTCCTGCTGCTGCTAGCATGTCTGAGTCTGAGTCTATATAAATTCG
N F K Y T G T T Q S Y R T Q L N N F K A -
AAAACTCCAAATAGTAGCGGGATATTACTGGAGGACTTTGTTACATGGTCAGCAGATGG
9181 +-----+ 9240
TTTTTGAGGTTTATCATCGGCCCTATATGACCTCCCTGAAACAATGTACCACTCGTCTCACC
K T P N S S G I L L E D F V T W S A D G -

PCT/US2005/027239

273/487

Figure 101V

9241 TGAACATAAGATTGTTCTGGAGATGGTGAAGTTATCAGATGTTTACGAAGAAACCTGT + 9300
-----+-----+-----+-----+-----+-----+
ACTTGTATTCTTAACAAGCACCTCTACCACTTTCATAATAGTCTACAAATGCTTCTTTGGACA
b E H K I V R G D G E S Y Q M F T K K P V -
9301 AACAGACCAATACGGAGTTTCATCAAAATACCTTCAATCACCTCCATGGAGCAGAGACTAA + 9360
-----+-----+-----+-----+-----+-----+
TTGTCTGGTTATGCCCTCAAGTAGTTTATGAAGTTAGTGGAGGTACCTCGTCTCTCGATT
b T D Q Y G V H Q I L S I T S M E Q R A K -
9361 ATTAGTTTCAGCGGGATATAGTTCTATGGAAGTACTGACTTGTATTATTTATTTGGCGTGATAG + 9420
-----+-----+-----+-----+-----+-----+
TAATCAAAGTCGCCCTATATCCAAAGATACCTTGACTGAACATAAAATAACCGCACTATC
b L V S A G Y R F Y G T D L Y L Y W R D S -
9421 TATTCTAGCCTATCCATTAACTCTAGTACCGATGGATTGATTACCAACCATGGTGACCTTAC + 9480
-----+-----+-----+-----+-----+-----+
ATAAGATCGGATAGGTAAATTGAGATCATGGCTAACCTAATGGTTGGTACCACCTGGGATG
b I L A Y P F N S S T D W I T N H G D P T -
9481 GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGTCTTCACCTGTTGGGT + 9540
-----+-----+-----+-----+-----+-----+
CTGAACCATATAATTGCCCTTTATACGGAGTCCCTACCGATACACAGAAGTGACAACCCCA
b T W Y Y N G N M A Q D G Y D V F T V G V -
9541 TGGTGTAAACGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTATTATGCAGAGCAT + 9600
-----+-----+-----+-----+-----+-----+
ACCACATTTGCCCTAGGACCATGCTACTTCTGTTGCCGATGATCTAAATACGTCCTCGTA
b G V N G D P G T D E A T A T R F M Q S I -
9601 CTCTAGTTCTCCTGACAACTACACTAACGTAGAGATCCATCTCAGATTTTACAGAATT + 9660
-----+-----+-----+-----+-----+-----+
GAGATCAAGAGGACTGTTGATGTGATTGCATCGTCTAGGTAGAGTCTAAATGTTCTTAA

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -
9661 GAATCGCTACTTCTATATCTGCTCAATGAGAAGAAATCTATCGAAATGGTACGATTAC 9720
CTTAGCGATGAAGATATGATAGCAGTTACTCTCTCTTTAGATAGCTTTACCATGCTAATG
N R Y F Y T I V N E K K S I E N G T I T -
9721 AGACCGGATGGGGAACCTAATTGATTTCCAAATGGGAGCAGATGGAGGTTTGTATCCAGC 9780
TCTGGGCTACCCACTTGATTAACCTAAAGGTTAACCCCTCGTCTACCTCCCAACATAGGTCG
D P M G E L I D F Q L G A D G R F D P A -
9781 GGATACACTTTAACTGCAAAACGATGGTAGTTCGTTGGTCAATAATCTCCTACTGGGGG 9840
CCTAATGTGAATTGACGTTTGTCTACCATCAAGCAACCACTTATTACAGGGATACCCCC
D Y T L T A N D G S S L V N N V P T G G -
9841 ACCACAAAATGATGGTGGCTTGTCTAAATAAAGTCAAAAGTCTCTATGATACGACTGAGAA 9900
TGGTGTCTTACTACCAACCGAAGCAATTTTACGTTTTCACAGATAGTACTGCTACTCTT
P Q N D G G L L K N A K V F Y D T T E K -
9901 AAGGATCGTGTACAGGTTTGTACCTTGGAAACGGGTGAAPAAAGTTACATTGACTTATAA 9960
TTCCTAAGCACATTTGTCCTCAACATGGAACCTTGCCCACTTTTTCATGTAACTGAATATT
R I R V T G L Y L G T G E K V T L T Y N -
9961 TGTTCGTTGAATGACCAATTTGTAAGCAATAAATCTATGACACGAAATGGTCGAACAAC 10020
ACAAGCGAACTTACTGGTTAAACATTCGTTATTAAAGATACTGTGCTTACCAGCTGTGTG
V R L N D Q F V S N K F Y D T N G R T T -
10021 CCTACACCCCTAAGGAAGTAGAAAAGAACACAGTCGCGGACTTCCCGATTCCCTAAGATTGCG 10080

Figure 101X

GGATGTGGGATTCCTTCATCTTTCTGTGTCAACGGCTGAAGGGCTAAGGATTCCTAAGC
L H P K E V E K N T V R D F P I P K I R -
10081
TGATGTACAAAGTATCCAGAAATCACAATTCCTCAAAAGAGAAAAACCTGGTGAATGA
ACTACATGCTTTCATAGGCTTTAGTCTTTAGGTAAAGTTTCTCTTTTGTGAACACCTTTAACT
D V R K Y P E I T I P K E K K L G E I E -
10141
GTTTATTAGATCAATAAGAAATGATAAAAAACCACTGAGAGATGCCGCTTTAGTCTTCA
CAAATAATCTAGTTATCTTACTATTCTTTTGGTACTCTCTACGCCAGAAATCAGAACT
F I K I N K N D K K P L R D A V F S L Q -
10201
AAAACAACATCCGGATTATCCAGATATTTATGGAGCTATTGATCAAAAATGGCACCTTATCA
TTTTGTTAGGCCCTAATAGGCTATAAAATACCCTCGATACTAGTTTACCGTGAATAGT
K Q H P D Y P D I Y G A I D Q N G T Y Q -
10261
AAATGTGACAAACAGGTGAAGATGGTAAGTTGACCTTTAAATCTGTCAATGGGAATA
TTTACACTCTGTGCCACTTCTACCAATCAACTGGAAATTTTATAGACAGCTACCCCTTAT
N V R T G E D G K L T F K N L S D G K Y -
10321
TCGATTATTGAAAAATCTGAACCACTGGTTATAAACCCGTCAAAATAGCCCTATCGT
AGCTAATAAACCTTTTAAGACTTGTGACCAATATTTGGGCAAGTTTATTCGGATAGCA
R L F E N S E P A G Y K P V Q N K P I V -
10381
TGCCCTCCAAATAGTAATGGAGAGTCAGAGATGTGACTTCAATCGTTCACAAAGATAT
ACGGAAGTTTATCATTTACCTCTTACCTCTCTACACTGAAGTTAGCAAGGTGTTCTATA
A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

10441 ACCAGCGGTTACGAGTTTACGAATGATAAGCACTATATATCACAATGAGCCAAATTCCTCC + 10500
-----+-----+-----+-----+-----+-----+-----+-----+
10501 TGGTCGCCCAATGCTCAAAATGCTTACTATTCGTGATATAGTGTITACTCCGGTTAAGGAGG
-----+-----+-----+-----+-----+-----+-----+-----+
b P A G Y E F T N D K H Y I T N E P I P P -
-----+-----+-----+-----+-----+-----+-----+-----+
10501 AAAAAAGAAATATCCTCGAACTGGTGGTATCGGAATGTGCCATTCTATCTCATAGGTTG
-----+-----+-----+-----+-----+-----+-----+-----+
TTTTTCTCTTATAGGAGCTTGACCACCATAGCCCTTACAACGGTAAAGATAGACTATATCCAAC
-----+-----+-----+-----+-----+-----+-----+-----+
b K R E Y P R T G G I G M L P F Y L I G C -
-----+-----+-----+-----+-----+-----+-----+-----+
10561 CATGATGATGGGAGGAGTTCTATATATATACACACGGAAACATCCGTAAAGTGTAGCAATGAG
-----+-----+-----+-----+-----+-----+-----+-----+
GTACTACTACCCCTCCTCAAGATAATATGTGTGCCCTTTGTAGGCATTTTCACATCGTTACTC
-----+-----+-----+-----+-----+-----+-----+-----+
b M M M G G V L L Y T R K H P *
-----+-----+-----+-----+-----+-----+-----+-----+
10621 AAATGATAATATCGATACCTCGAGCGGATATTTTAAAGTAGTACACTCAAGAGAGATTT
-----+-----+-----+-----+-----+-----+-----+-----+
TTTACTATTATAGCTATGAGACTCGCTATGAAATTTCTTCATCGTGAGTTCTCTCTCTAAA
-----+-----+-----+-----+-----+-----+-----+-----+
10681 AAGTTTACTTGGTGAACACAGTTTCTTCGCCAAGTAAACACCATTGAAAGGGGAGATG
-----+-----+-----+-----+-----+-----+-----+-----+
TTCAAAATGAACCACTTTTGTCAAAGAGAGCGGTTCAATTTGGTGGTAACTTTCCCTCTCTAC
-----+-----+-----+-----+-----+-----+-----+-----+
10741 TTTTCGAAACTTCACAGAAAAAGGATTATTATTGTCATGTGTAATTCATTACATTGC
-----+-----+-----+-----+-----+-----+-----+-----+
AAAAGCTTTTGAACGTGCTTTTTTCCATAATAATAACAGTACACATTAAAGTAATGTAACG
-----+-----+-----+-----+-----+-----+-----+-----+
10801 TCACAGTTGATTTTAAGAGATATGAATAAGGAGAAATCATGAATCAATCAACAAATTTT
-----+-----+-----+-----+-----+-----+-----+-----+
AGTGCAACTAAATTCCTCTATACTTATTCTCTCTTTTACTACTTTTACTAGTTGTTTAAAA
-----+-----+-----+-----+-----+-----+-----+-----+
c M K S I N K F L - orf4_670, homologue of sp0463, LPXTG
-----+-----+-----+-----+-----+-----+-----+-----+
TAACAATGCTTGCTGCCCTTATATCTGACACGGAGTAGCCTGTTTTCAGCTGCAACAGTTT

[illegible]

PCT/US05/27239

278/487

Figure 101AA

```
11281 TTGGTCCTTAATGGGCAAGTATTAAACAGTTCAAAAGCCGTACCTGCTCTTGTAACTCTTC + 11340
-----+-----+-----+-----+-----+-----+-----+
AACCAGGATTACCCGTTCAATTAATTGTCCAAGTTTTCGGCATGGACGAGAACATTGAGAAG
      G P N G Q V L T G S K A V P A L V T L P -
11341 CACTTGTTAACAATAATGGTACAGTAATTGATGCACATGTTTCCCTAAATAATTCATATA + 11400
-----+-----+-----+-----+-----+-----+-----+
GTGAACAATTGTTATTACCATGTCTATTAACTACGTGACAAAAGGGATTTTAAAGTATAT
      L V N N N G T V I D A H V F P K N S Y N -
11401 ATAAACCAGTTGTAGATAAAAAGAATTGCTGATACCTTTGAATTATAACGATCAAAATGGTC + 11460
-----+-----+-----+-----+-----+-----+-----+
TATTGGTCAACATCTATTTTCTTAACGACTATGAACCTTAATATGCTAGTTTACCAG
      K P V V D K R I A D T L N Y N D Q N G L -
11461 TGTCTATCGGTACTAAAATCCCATATGTTGTTAATACAAACAATTCCAAGTAATGCAACAT + 11520
-----+-----+-----+-----+-----+-----+-----+
ACAGATAGCCATGATTTTGGGTATACAACAATTATGTTGTTAAGGTTCAATACGTTGTA
      S I G T K I P Y V V N T T I P S N A T F -
11521 TTGCAACTTCATTTTGGTCAGATGAATGACAGAGGTCTAACTTATATAATGAAGATGTAA + 11580
-----+-----+-----+-----+-----+-----+-----+
AACGTTGAAGTAAACCCAGCTCTACTTTACTGTCTTCCAGATTGAATATTACTTCTACATT
      A T S F W S D E M T E G L T Y N E D V T -
11581 CAATTACTTTGAATAATGTAGCTATGGATCAAGCTGATTATGAGTCACATAAGGAATA + 11640
-----+-----+-----+-----+-----+-----+-----+
GTTAATGAACCTTATTACATCGATACCTAGTTCGACTAATACTTCAGTGATTTCCTTTAT
      I T L N N V A M D Q A D Y E V T K G N N -
11641 ATGGCTTTAACTTAAATTAACAGACGAGTTTAGCTAAATAATTAATGGAAGGATGCAG + 11700
-----+-----+-----+-----+-----+-----+-----+
TACCGAATTGAATTTTAAATTGCTTCGTCCAAATCGATTTTAAATTACCATTCTACGTC
```

Figure 101AB

C	G	F	N	L	K	L	T	E	A	G	L	A	K	I	N	G	K	D	A	D	-
11701	ACCAAAAA	TCCAAAT	TACTT	ACTCAG	CTACTT	TGAACT	CACTT	TGAACT	CACTT	GTCTG	TGTCAG	ACATTC	-----	+	-----	+	-----	+	-----	+	11760
	TCGTTTTT	TAGTTT	TAATGA	TAGT	CGATGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA	ACCTG	AGTGA
C	Q	K	I	Q	I	T	Y	S	A	T	L	N	S	L	A	V	A	D	I	P	-
11761	CTGAAAGT	AAAGCAT	ATTAC	ATATCA	TATAC	TACG	GAATCA	TGGA	AAATCA	TGGA	AAATCA	TGGA	AAATCA	TGGA	AAATCA	TGGA	AAATCA	TGGA	AAATCA	TGGA	AAATCA
	GACTTTCA	TTCGTAT	ATAATG	TATAG	TAAATG	CCCTTT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT	AGTTCT
C	E	S	N	D	I	T	Y	H	Y	G	N	H	Q	D	H	G	N	T	P	K	-
11821	AACCAACT	AAACCT	AAATAA	TGGTCA	AAATTA	CAGTAA	CTAAAG	ACATGA	GGACAG	TCAACCTG	-----	+	-----	+	-----	+	-----	+	-----	+	11880
	TTGGTTGA	TTTGGAT	TATATAC	CGATTAA	TGTCAT	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC	TGATTC
C	P	T	K	P	N	G	Q	I	T	V	T	K	T	W	D	S	Q	P	A	-	
11881	CTCCTGAG	GGGTGA	AAAGG	ACGTG	TTCAACT	TGTA	AAATG	CCAAAG	ACTGGT	GAGAAA	GTGG	-----	+	-----	+	-----	+	-----	+	-----	+
	GAGGACTC	CCCCACT	TTTCGG	TGACAA	GTGTA	CAATTA	CGGTT	CTGAC	CACTCT	TTTCAG	C-----	+	-----	+	-----	+	-----	+	-----	+	11940
C	P	E	G	V	K	A	T	V	Q	L	V	N	A	K	T	G	E	K	V	G	-
11941	GTGTCCTG	TAGAAAC	TTCAG	AAAAA	TAAATG	GACAT	ATAC	TACTTG	GAGTGG	TCTAG	ATAA	TT-----	+	-----	+	-----	+	-----	+	-----	+
	CACGAGGA	CATCTTG	AAAGTC	TTTTT	TATTA	ACCTG	TATGA	AACTC	ACCA	GTCTAT	TAA-----	+	-----	+	-----	+	-----	+	-----	+	12000
C	A	P	V	E	L	S	E	N	N	W	T	Y	T	W	S	L	D	N	S	-	
12001	CTATTGA	ATACAAA	GTGAA	GAAG	ATAA	TGATG	ATCT	CAGCTG	AAATAC	ACAGT	AGAG	-----	+	-----	+	-----	+	-----	+	-----	+
	GATAACTT	ATGTTT	CAACTT	CTCTT	CTTAT	TATAC	CTATG	AGTCG	ACTTAT	GTGTC	ATCTCT	-----	+	-----	+	-----	+	-----	+	-----	+
C	I	E	Y	K	V	E	E	Y	N	G	Y	S	A	E	Y	T	V	E	S	-	
12061	GCAAAAGG	GAAGTT	GGGGT	TAAAAA	ACTGG	GAAG	ATAA	TAACTC	CAATCA	ATCCCTG	-----	+	-----	+	-----	+	-----	+	-----	+	12120

CGTTTCCCTTCAACCCCAATTTTGTACCTTTCTAATTATTTGGGTCGAGGTAGTTAGGAC

12121
AAGAACCCAGGTGTAAAAACATACGGTAAAGTCTGTCAAAGTAGACCAAAAAGATATCTC
-----+-----+-----+-----+-----+-----+-----+
12180
TTCTTGTGTGCACATTTTGTGTATGCCATTTTTCAAACAGTTTCTACTGGTTTTTCTATGAG

12181 GTCTAGAAAATGCCGACGTTGGTTGTTAAAAAAGCAGATAGCAATAAATATATATGCCCTTTA
-----+-----+-----+-----+-----+-----+-----+
12240 CAGATCTTTTACGCGTCAAGCAACAATTTTTTCGCTCTATCGTTATTTTATATAACCGGAAT

AGTCAACTGCACAACAAGCTGCAGATGAAAAAGCAGCAGCAACTGCAAAACAAAAATTGG
-----+-----+-----+-----+-----+-----+-----+
12241
TCAGTTCACGGTGTGTCGACGCTACACTTTTTCGTCGTCGTTGACGCTTTGCTTTTAAAC

ATGCAGCGGTAGCAGCTTACACAATGCTGCAGATAAGCAAGCGCTCAAGCTCTAGTAG
12301-----+-----+-----+-----+-----+-----+-----+ 12360
TACGTGGCCATCGTTCGATATGTTTATCGACGCTCATTCGTTCCGCGAGTTCGAGATCATC

ATCAAGCAGCAGCAAGAAATACAATGTAGCTTACAAGAAGCCAAATTTGGTTATGTTGAAG
 12361 -----+-----+-----+-----+-----+-----+-----+-----+
 TAGTTCCGTCGTTCTTATGTTTACATCGAAATGTTTCTTCGGTTTAAACCAATACAACACTTC
 12420 -----+-----+-----+-----+-----+-----+-----+-----+

12421 TAGCTGGAAAAGATGAAGCAATGGTTCTTACTTCTAAATACGGATGGTCAATTCCAAATTT
-----+-----+-----+-----+-----+-----+-----+
ATCGACCTTTTCTACTCTCGTTACCAAGAATGAAGATTATGCCCTACCGATTAGGTTTAAA 12480

AGKDEAMVLTSENFGDGOHSA

12481

CAGGTCCTTGCCTGGTACCTTATATAATTAGAAGAAATTAAAGCTCCAGAAGGTTTTCGCA
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTCCAGAACGACGACCATGATGATATTTATCTTCTTTAATTTGAGGTCCTTCACAAAACGGT
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 12540

GLAAGT YKKLE I K A P E G F A K -

AAATTGATGATGTTAGTAATTTGTTGTTGGAGCAGGTTCTTTGGAATCAAGGTGAGTTTAATT
 12541
 TTTTAACACTACATCTTTAAACAACAACCTCGTCCAGAAGAACCTTAGTTCCACTCAAAATTAA
 12600

C I D D V E V V G A G S W N Q G E F N Y -

12601
ACTTAAAGAGTGTTCAAAAGATGAGCGCTACAAAGTAGTGTCAACAAAAAATCACTATCCC
-----+-----+-----+-----+-----+-----+
12660
TGAATTTTCTACAAGTTTTCCTTACGCGGATGTTTTCATCATGTTGTTTTTTTATGTGATAGG

C L K D V Q K N D A. T K V V N K K I T I P -

12661 CACAAACGGGTGGTATTGGTACAAATTATCTTTGCTGTAGCGGGGGCTGCGATTATGGGPA 12720
GTGTTTGCCCAACCATTAACCATGTTTAATAGAAACGACATCGCCCCCGAGGTTAATAATACCCAT

C Q T G G I G T I I F A V A G A I M G I -

12721
TTGCAGTGTACGCATATGTTAAAAACAACAAGATGAGGATCAACTTGCCTTAAGTAAGAG
-----+-----+-----+-----+-----+-----+
AAGTGCACATGCGTATACAAATTTTCTGTGTTTCTACTCTCCTAGTTGAAACGAATTCATTCTC
12780

A V Y A Y V X N N K D E D Q L A *

12781 AGAAAGGAGCCATTGATGACAAATGCAGAGAAAATGCAGAGAAAATGATTAGTGCGTATCTCTTTT
-----+-----+-----+-----+-----+-----+-----+-----+ 12840
TCTTTCTCTGGGTAACATCTGTTACGTCCTTTTACGTCCTTTTACATAATCAGCATAGAAGAAA

a a M T M Q K M Q K M I S R I F F -orf5_670, homologue of sp0464, LPXTG

12841 GTTATGGCTGTGTTTTTCTCTTGTATGGGGTGCACATGCAGTCCCAAGCGCAAGAAGAT
-----+-----+-----+-----+-----+-----+-----+
CAATACCGAGACACAAAAGAGAGAACATACCCCACTGTACGTCAGGTTCCGCTTCTTCTA

PCT/US05/27239 282/487

Figure 101AE

a V M A L C F S L V W G A H A V Q A Q E D -
CACACGTTGGTCTTGGCAATTGGAGAACTATCAGGAGGTGGTTAGTCAATTGCCATCTCGT 12901
GTGCGCAACGAGAACGTTAACTCTTGTAGTCTCTCCACCAATCAGTTAACGGTAGAGCA 12960
a H T L V L Q L E N Y Q E V V S Q L P S R -
GATGGTCATCGGTTGCAAGTATGGAAGTTGGATGATTCGTATTCTATGATCATCGGGTG 12961
CTACAGTAGCCACGTTTCATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCCCAC 13020
a D G H R L Q V W K L D D S Y S Y D D R V -
CAAAATTGTAAGAGACTTGCCATTGCGGATGAGAAATAAAGTTCTTCTTCAAAAAGACT 13021
GTTTAACATTCTCTGAACGTAAGCACCCCTACTCTTATTGAAAGAGAAAGTTTCTTGA
a Q I V R D L H S W D E N K L S S F K K T -
TCGTTGAGATGACCTTCCTTGAGATCAGATGGAATGATATCATATCCAAATGGTCTT 13081
AGCAAACTCTACTGGAAGGAACCTCTTAGTCTAATTCATAGAGATTAAGTTTACCAGAA 13140
a S F E M T F L E N Q I E V S H I P N G L -
TACTATGTTGGTCTATTATCCAGACGGATGCGGTTTCTTATCCAGCTGAATTCCTTTT 13141
ATGATACAAGCGAGATAATAGGTCTGCTACGCCCAAGAATAGGTGACTTAAAGAAAAA
a Y Y V R S I I Q T D A V S Y P A E F L F -
GAAATGACAGATCAACGGTAGAGCCTTTGGTCAATGTAGCGAAAAAACAGATACAATG 13201
CTTTACTGTCTAGTTTGCCATCTCGGAACCAGTAAATCGCTTTTGTCTATGTTAC 13260
a E M T D Q T V E P L V I V A K K T D T M -
ACACAAAGGTGAAGCTGATAAAGGTGGATCAAGACCAATCGTTGGAGGTGTCCGGC 13261
13320

PCT/US05/27239 283/487

Figure 101AF

13321 TGTGTTTCCACTTCGACTATTTCACCTAGTTCTGGTGTAGCGAACCTCCACAGCCG
a T T K V K L I K V D Q D H N R L E G V G -
TTTAAATTGGTATCAGTAGCAAGAGATGGTCTCTGAAAAAGAGGTTCCCTTGATTGGAGAA
13380
AAATTTAACCATAGTCATGGTTCTCTACCAAGACTTTTTCCTCAAGGGAACATAACCTCTT
a F K L V S V A R D G S E K E V P L I G E -
TACCGTTACAGTTCCTCTGGTCAAGTAGGGAGAGACTCTCTATACTATGATAAAAATGGAGAG
13381
ATGGCAATGTCAGAGACCAAGTTCATCCCTCTCTGAGAGATGACTATTTTACCTCTC
a Y R Y S S S G Q V G R T L Y T D K N G E -
ATTTTGTGACAAATCTCTCTTGGGAACATATCGTTTCAAGGAGGTGGAGCCACTGGCA
13441
TAAACACTGTTTAGAAGGAGAACCTTGATAGCAAGTCTCCACCTCGGTGACCGT
a I F V T N L P L G N Y R F K E V E P L A -
GGTATGCTGTACGACGCTGGATACGGATGTCAGCTGGTAGATCATCAGCTGGTGACG
13501
CCGATACGACAATGCTGGACCTATGCTACAGTCCGACCATCTAGTAGTCGACCACTGC
a G Y A V T T L D T D V Q L V D H Q L V T -
ATTACGGTGTCAATCAGAAATTACCACGTTGGCAATGTTGACTTTATGAAGGTGGATGGT
13561
TAATGCCAACAGTTAGTCTTTAATGGTGACCGTTACACTGAAATACTTCCACCTACCA
a I T V V N Q K L P R G N V D F M K V D G -
CGGACCAATACCTCTCTTCAAGGGCAATGTTCAAGTCAATGAAGAAGAAAGCGGACAC
13621
GCCTGGTATGGAGAGAAGTTCCCGTTACAAAGTTTCAGTACTTTCTCTTTCGCGCTGTG
a R T N T S L Q G A M F K V M K E E S G H -

13681	a	TATACTCCTGTTCTTCAAAATGGTAAAGGAGTAGTGTCTTAACATCAGGGAAGATGGTCTGT -----+-----+-----+-----+-----+-----+-----+-----+ ATATGAGGACAAAGAAAGTTTACCATTTCCTTCATCAACATTTGTAGTCCCTTTCTTACCAGCA	13740
		Y T P V L Q N G K E V V V T S G K D G R -	
13741		TTCCGAGTGGAAAGGTCTAGAGTATGGACATACATAATTTATGGAGCTCCAAGCTCCAACCT -----+-----+-----+-----+-----+-----+-----+-----+ AAGGTCACCTTCAGATCTCATACCTGTATGATAAATACCTCGAGGTTCCGAGGTTGA	13800
	a	F R V E G L E Y G T Y Y L W E L Q A P T -	
13801		GGTTATGTTCAANTTAACATCGCCTGTTTCCTTTTACAATCGGGAAGATACCTCGTAAGGAA -----+-----+-----+-----+-----+-----+-----+-----+ CCAATACAAGTTAATTGTAGCGGACAAAGGAAATGTTAGCCCTTTCTATGAGCATTCCTT	13860
	a	G Y V Q L T S P V S F T I G K D T R K E -	
13861		CTGCTAACAGTGGTTAAATAACAAGCGACCAAGGATGTATGCCAGATACAGGGGAA -----+-----+-----+-----+-----+-----+-----+-----+ GACCATGTCAACCAATTTTATTTGTTCTCGTGGTGCCTAACTACAGGTCTATGTCCCTT	13920
	a	L V T V V K N N K R P R I D V P D T G E -	
13921		GAAACCTTGATATCTTGATGCTTGTTCGCATTTTGTGTTTGGTAGTGGTTATTATCTT -----+-----+-----+-----+-----+-----+-----+-----+ CTTTGGAACATATAGAACATACGAACAACGGTAAACCAACAACCATCACCAATAATAGAA	13980
	a	E T L Y I L M L V A I L L F G S G Y Y L -	
13981		ACGAAAAACCAATAACTGATATTCAATGTACATCATATTATGAAAAAGATAGCAGGCTGA -----+-----+-----+-----+-----+-----+-----+-----+ TGCTTTTITGGTTTATTGACTATAAGTTACATGATGTAGTAATACITTTTTTCTATCGTCCGACT	14040
	a	T K K P N N *	
14041		AGGGAAGACCAGAGTACTCTGAGGTGATGTTAATCAGGAATCATGGTGTATGGGCATGAA -----+-----+-----+-----+-----+-----+-----+-----+ TCCTTCTGCTCATGAGACTCCACTACAATTAGTCCTTAGTACCCTACACCGTACTT	14100

Figure 101AH

14101 TCACAATAACGGATATAGGCTGGCAGATGTGCCAGCCTCATGTGGTATTGTTTG + 14160
-----+-----+
AGTGTATTGCTATATACCTCGACCCGCTCTAACACGGTGGAGTAACACCCCAATAACAAAC
14161 TAAAAAGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGACTGGGATTCGTATT + 14220
-----+-----+
ATTTTGTATCTCTGACACAGACCATTAAGTAAATCCTTACCTGCTCTGACCCCTAAGACTAA
14221 TAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT + 14280
-----+-----+
ATTTTACCTACCACTTAGTCTTCTTTTACTCTCTAAAGAGCAAGAGAAATCCTCTATCCTA
14281 TGTCTGTTAGGAAAAGCGATATAAATGATGAGTTTGAAGATAAAGGGATGCTGATATAAAAT + 14340
-----+-----+
ACAGACAAATCCTTTTTCGCTATTTTACTACTCAAACTCTATTTTCCCTACGACTATTTTAA
M L I K M -orf6_670, homologue of sp0466, sortase
CGTAAAAAGAAAAAGCAAAAACGAAATAATCTCCTATTAGGAGTGGTATTTTTCATTGG + 14400
-----+-----+
CCATTTTGTGTTTTTTCGTTTTTTCCTTTATTAGAGGATAATCCTCACCATAAAAAGTAACC
V K T K K Q K R N N L L L G V V F F I G -
14401 AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGTCTGTTATATCGAGTGGATCAAAATCA + 14460
-----+-----+
TTACCGCCATACCGCATAGCGGACCCACAGAGCGGAACATAATAGCTCACCCTAGTTTAGT
M A V M A Y P L V S R L Y Y R V E S N Q -
14461 ACAAAATTGCTACTTTGATAAGGAAAAACCAACGTTGGATGAGGCTGACATGATGAACG + 14520
-----+-----+
TGTTTAACGACTGAACCTATTCCTTTTTCGTTGCAACCTACTCCGACTGTAACACTACTTGC
Q I A D F D K E K A T L D E A D I D E R -
AATGAAATGGCACAAGCCTTCAATGACTCTTTTGAATAAATGATGAGTGGCGATCCTTG

14521	----	TTACTTTAA	CCG	TGTT	CGGAAG	TACT	CTG	AGAAA	CTTT	ATTACAT	CAT	CAC	CGCTAG	GAAC	14580								
b		M	K	L	A	Q	A	F	N	D	S	L	N	V	V	S	G	D	P	W	-		
14581		GT	CGGA	AGAA	ATGA	AGAAAA	AGGG	CGAG	CAG	AGTAG	TGC	ACGT	ATGT	TAG	AAAT	CCAT	GA	14640					
		CAG	CGT	CTCT	TTT	AGTT	CTTT	TTT	CCCG	CTCT	CTCAT	AGT	GCAT	ACA	ATCT	TTAG	GTACT						
b		S	E	E	M	K	K	K	G	R	A	E	Y	A	R	M	L	E	I	H	E	-	
14641		GG	CGAT	GGG	GCAT	GTG	GAAT	CCCG	TAT	TGAC	GTG	GAAT	TGC	CGT	TTAT	TGCT	GGTAC	14700					
		CG	CTAC	CCCC	GTAC	ACCT	TTT	AGG	GGCA	TAA	CTG	CACT	AAAC	GGCC	AAAT	AGCAC	CATG						
b		R	M	G	H	V	E	I	P	V	I	D	V	D	L	P	V	Y	A	G	T	-	
14701		TG	CTGA	AGAG	GTAT	TGC	AGCA	AGGG	GCTG	GGCAT	CTAG	AGG	GAAC	TCT	CTG	CGGAT	CGG	14760					
		AC	GA	CTT	CTCCA	TAA	ACG	TGTT	CCCC	CGAC	CCGT	AGAT	CTCC	TTGA	AGAG	ACGG	CTAG	CGC					
b		A	E	E	V	L	Q	G	A	G	H	L	E	G	T	S	L	P	I	G	-		
14761		AG	GC	AA	TT	CGAC	CCAT	CGG	TGAT	TAC	GGCAC	ATAC	AGT	TTT	GCC	AA	CAGC	AA	AGAT	GT	14820		
		TCC	GT	TA	AG	CTGG	GTAC	CGCC	ACT	AAAT	GCC	GTG	TAT	GTCC	AAAC	GGT	TGTC	GAT	CT	TAC	AA		
b		G	N	S	T	H	A	V	I	T	A	H	T	G	L	P	T	A	K	M	F	-	
14821		TAC	GG	AA	TTT	GAC	CAAA	CTTAA	AGT	TGG	GAATA	AGT	TTT	TAT	TGT	GC	ACA	TAT	CA	AGG	AA	GT	14880
		AT	GC	CTAA	AACT	GGTT	GAAT	TTCA	ACCC	CTAT	TCA	AAAT	ACAC	CGT	GT	TAT	AGT	TC	CT	TC	TCA		
b		T	D	L	T	K	L	K	V	G	D	K	F	Y	V	H	N	I	K	E	V	-	
14881		GAT	GG	CTAT	CA	AGT	GGAT	CA	AGTAA	AGG	TGAT	TAG	CGG	CAG	CACT	TTT	TGAT	GTAT	TAT			14940	
		CT	AC	CGGAT	AGT	TAC	CTAG	TTCA	TTT	CTC	ACTA	CTCT	CGCT	GC	TTG	AAAC	TACT	AA	ATAA				
b		M	A	Y	O	V	D	O	V	K	V	I	E	P	T	N	F	D	D	L	L	-	

287/487

Figure 101AJ

14941 GATTGTACAGGTCATGATTATGTGACCTTGTGACTTGTAGCCCATCATGATCAATAC 15000
CTAACATGGTCCAGTACTAATACACTGGAACGACTGAACATGCGGTATGTACTAGTTATG
b I V P G H D Y V T L L T C T P Y M I N T -
15001 CCATCGTCTATTGTTTCGGGGCATCGGATACCGTACGTACGAGAGGTGAGGAAGAATT 15060
GGTAGCAGATAACCAAGCCCCCGTAGCCTATGGCATGCACTGCTCTCCAACTCCTTCTTAA
b H R L L V R G H R I P Y V A E V E E F -
15061 TATTGCAGCAACAACTCAGTCACTCTATCGCTACCTGTTTATGTGGCAGTTGGTTT 15120
ATAACGTCGTTGTTGAGTCAGTAGAGATAGCGATGGACAAAATACACCGTCAACCAAA
b I A A N K L S H L Y R Y L F Y V A V G L -
15121 GATTGTGATCTTTTATGGATTATTCGACGCTTTCGCGAAGAGAAACAAACCGGAAAA 15180
CTAACACTAAGAAAATACCTAATAAGCTGCGAAGCGGTTCTCTTTTGTGGCTTTT
b I V I L L W I I R R L R K K K Q P E K -
15181 GGCTTTGAAGGGCGCTGAAAGCAGCAAGGAAGGAAGTGAAGTGGAGGATGGACAACAGTA 15240
CCGAAACTCCGGGACTTTCGTGCTCCTTCCCTTCACTTCCACCTCCCTGTTGTCAT
b A L K A L K A A R K E V K V E D G Q Q * -
15241 GACGTTACGAAAAAAGGCACAAAAAGAAAGAACATCCGCTGATCCCTTCTTCTGATTT 15300
CTGCAAGTCTTTTTCGGTGTGTTTCTTCTTGTAGGCGACTAGGAAGAAGACTAAA
TCTTAGTAGGATTCGGCGTTGGGATATATCCATTGGTGTCTGTTATTATTATTCGTATTG
15301 AGAATCATCTAAGCGCAAGCGCTATATAGGTAACCAAGCAATATAATAGCATTAAC 15360
V S R Y Y R I E -orf7_670, homologue of sp0467, sortase

PCT/US05/27239

288/487

Figure 101AK

```
15361 AGTCAAACGAGGTTATTAAAGAGTTTGATGAGACGGTTTCCAGATGGATAAGGCAGAAC 15420
-----+-----+-----+-----+-----+-----+-----+-----+
TCAGTTTGCTCCCAATAATTTCCTCAAACTACTCTGCGCAAGGGTCTACCTATTCGGTCTTTG
C S N E V I K E F D E T V S Q M D K A E L -
15421 TTGAGGAGCGTTGGCGCTTGGCTCAAGCCTTCAATGCGACCTTGAACCATCTGAAATTC 15480
-----+-----+-----+-----+-----+-----+-----+-----+
AACTCCTCGAACC CGGAACCGAGTTCCGGAAGTTACGCTGGAACCTTGGTAGACTTTTAAG
C E E R W R L A Q A F N A T L K P S E I L -
15481 TTGATCCTTTTACAGAGCAAGAGAAAAAGAAAGCGCTCTCAGAAATATGCCAATATGCTAA 15540
-----+-----+-----+-----+-----+-----+-----+-----+
AACTAGGAAAAATGTCTCGTTCGTTCTCTTTTCTTTCCGCGAGAGTCTTATACGGTTATACGATT
C D P F T E Q E K K K G V S E Y A N M L K -
15541 AGGTCCATGAGCGGATTGGCTATGTGGAATTCCTCGGATTGATCAGGAAATTCGGATGT 15600
-----+-----+-----+-----+-----+-----+-----+-----+
TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAACTAGTCTCTTTAAGGCTACA
C V H E R I G Y V E I P A I D Q E I P M Y -
15601 ATGTCCGAACGAGTGAGGAAATTCCTTCAGAAAGGCGCGAGGATTGCTAGAGGGAGCTTCGT 15660
-----+-----+-----+-----+-----+-----+-----+-----+
TACAGCCTTGCTCAGTCTCTTTAAGAAGTCTCCCGGCTTAACGATCTCCCTCGAAGCA
C V G T S E E I L Q K G A G L L E G A S L -
15661 TACCGGTTGGTGGTGAATAATACCCACACAGTTGTCTACTGCTCATAGAGGATTACCGACGG 15720
-----+-----+-----+-----+-----+-----+-----+-----+
ATGGCCAACCACCACTTTTATGGGTGTGTCAACAGTGACGAGTATCTCCTAATGGCTGCC
C P V G G E N T H T V V T A H R G L P T A -
15721 CAGAACTGTTTAGTCAATTGGATAAGATGAAAAAGGGGATGCTTTTATCTTCACGTTT 15780
-----+-----+-----+-----+-----+-----+-----+-----+
GTCCTTGACAAATCAGTTAACCTATTCTACTTTTTTCCCTTACAGAAAAATAGAAAGTGCAAA
```

PCT/US05/27239

289/487

Figure 101A

```
c      E L F S Q L D K K K G D V F Y L H V L -
15781 TAGACCAGGTGTTGGCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG
-----+-----+-----+-----+-----+ 15840
ATCTGGTCACAACCGGATGGTTCACCTAGTCTATAAAGCTGCCAACTCGGTTTACTGAAAC
-----+-----+-----+-----+-----+

c      D Q V L A Y Q V D Q I L T V E P N D F E -
15841 AGCCTGTCTTGATTCAACATGGGGAGATTATGCGACCTTGTGTGACCTGTACACCGTATA
-----+-----+-----+-----+-----+ 15900
TCGGACAGAACTAAGTTGTACCCCTTCTAATAGCTGGAACAACCTGGACATGTGGCATAT
-----+-----+-----+-----+-----+

c      P V L I Q H G E D Y A T L L T C T P Y M -
15901 TGATTACAGTCATCGTCTGTGTTGGTACGTTGGGAGCGGATTCCGTATACGGCACCAATTG
-----+-----+-----+-----+-----+ 15960
ACTAATTGTCAGTAGCAGACAACCATGCACCTTCCCTAGGCATATGCCGTGGTTAAC
-----+-----+-----+-----+-----+

c      I N S H R L L V R G K R I P Y T A P I A -
15961 CAGAGCGAATCGAGCGGTGAGAGCGGTGGGAAATTCGTGTGTGTTATTGCTAGCGG
-----+-----+-----+-----+-----+ 16020
GTCTCGCTTAGCTCGGCACCTCTCTGACCCGTTAAGACCAACCAATAACGATGCC
-----+-----+-----+-----+-----+

c      E R N R A V R E R G Q F W L L L L A A -
16021 CGTTGGTTATGATTCCTGGTATTGAGTTACGGGGGTATCGTCATCGTCCGATTTGTCAAAG
-----+-----+-----+-----+-----+ 16080
GCAACCAATACTAAGACATAACTCAATGCCCAACATAGCAGTAGCAGCGTAACAGTTTC
-----+-----+-----+-----+-----+

c      L V M I L V L S Y G V Y R H R R I V K G -
16081 GGCTAGAAAACAAATTGGAGGAGCATCATGTCAAGGCTAAGCTACAGAAATTACTAGGG
-----+-----+-----+-----+-----+ 16140
CCGATCTTTTGTAACTCCTCGTAGTACAGTTCCGATTCGATGTCCTTAATGATCCC
-----+-----+-----+-----+-----+

a      M S K A K L Q K L L G -orf8_670, homologue of sp0468, sortase
c      L E K Q L E E H H V K G *
```

290/487

Figure 101AM

16141 TATTGCTGATGCTGGTACCATTTGGTATTCCTGTTTATTTGTTTGGGAGATGGTGTTA 16200
-----+-----+-----+-----+-----+-----+-----+
ATAAACGACTACGACCATCGTAACCACTAAGGACAATAACAAAACCCGCTCTACCACAAT
a Y L L M L V A L V I P V Y C F G Q M V L -

16201 CAGTCTTTAGGACAAAGTAAAGTCAATGAGATATTTTCAGAAATCTGTGACGGCCGACAGT 16260
-----+-----+-----+-----+-----+-----+-----+
GTCAGAAATCCTGTTTCATTTTCCAGTACTCTATAAAAGTCTTAGACACATGCCCGGCTGTCA
a Q S L G Q V K G H E I F S E S V T A D S -

16261 TACCAAGACCAATTGCAACGGTCGTTGATTACAAATCAACGCTGGATTGCGCAAAATCGT 16320
-----+-----+-----+-----+-----+-----+-----+
ATGGTTCCTGTTAACGTTGCCAGCGAACTAATGTTAGTTGCGAACCTAAGCGTTTATAGCA
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -

16321 ATTGTAGATCCCTTTTGGCGGAAGGTATGAGTAAATACCAAGTCTCTGACGATCCT 16380
-----+-----+-----+-----+-----+-----+-----+
TAACATCTAGGAAAAAACCCGCTTCCCATCTCCATTTAATGGTTTCACAGACTGCTAGGA
a I V D P F L A E G Y E V N Y Q V S D D P -

16381 GATGCAGTCTACGGCTATTTGTCGATTCGAGTTGGAATCATGAGCCAGTTTATCTA 16440
-----+-----+-----+-----+-----+-----+-----+
CTACGTCAGATGCCGATAACAGCTAAGGCTCAACCTTTAGTACCTTCGGTCAAAATAGAT
a D A V Y G Y L S I P S L E I M E P V Y L -

16441 GGAGCGGATTACCATCATTTAGCAATGGGTTGGCCCATGTTGGATGGACGCCCTTCCT 16500
-----+-----+-----+-----+-----+-----+-----+
CCTCGCTAATGTTAGTAATTCGTTACCCCAACCGGGTACACCTACCTCGCGGAGAAGGA
a G A D Y H H L A M G L A H V D G T P L P -

16501 GTTGAGGGAAGGATTCGTTTCACTGATTGTTGGGACCGCTGCAGAACCAAGCCATGTC 16560
-----+-----+-----+-----+-----+-----+-----+
CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCCGTGGCAGCTCTTGGTTCCGGTACAG

Figure 101AM

a V E G K G I R S V I A G H R A E P S H V -
TTTTTCGCCCATTTGGATCAGCTAAAGTTGGAGATGCTCTTTATATGATAATGGCCAG 16561
-----+-----+-----+-----+-----+
AAAAGCGGTAAACCTAGTCGATTTTCAACCTCTACGAGAAATAACTATTACCGGTC
F F R H L D Q L K V G D A L Y Y D N G Q -
GAAATTGTAGAAATCATCAGATGATGGACACAGAGATTATTTACCGTCGGAATGGGAAAA 16621
-----+-----+-----+-----+-----+
CTTTAACATCTTATAGTCTACTACCTGTGCTCTTAATAAAATGGCAGCCTTACCCCTTTT
E I V E Y Q M M D T E I I L P S E W E K -
TTAGATCGGTTAGCTCTAAAAATATCATGACCTTGATAACCTGGATCCGATTCCCTACC 16681
-----+-----+-----+-----+-----+
AATCTAGCCAAATCGAGATTTTATAGTACTGGAACHTATTGGACGCTAGGCTAAGGATGG
L E S V S S K N I M T L I T C D P I P T -
TTTAAATAACGCTTATTAGTGAATTTTGAACGAGTCGCTGTTTATCAAAAAATCAGATCCA 16741
-----+-----+-----+-----+-----+
AAATTATTGGGAATAATCATTAAACCTTCTCAGCAGACAAATAGTTTATTAGTCTAGGT
F N K R L L V N F E R V A V Y Q K S D P -
CAACAGCTGCAGTTCGAGGTTGCTTTACGAAAGAGGACAACTGTATCGCGTGT 16801
-----+-----+-----+-----+-----+
GTTTGTGACGTCAACGCTCCCAACGAAATGCTTTCTCTGTTAGACATAGCGCACAA
Q T A A V A R V A F T K E G Q S V S R V -
GCAACTCTCAATGGTTGTACCGTGGCTAGTGGTACTGGCATTTCTGGGAATCCTGTTT 16861
-----+-----+-----+-----+-----+
CGTTGGAGAGTTACCAACATGGCAGCCGATCACCATTGACCGTAAAGACCCCTTAGGACAAA
A T S Q W L Y R G L V V L A F L G I L F -
GTTTGTGGAAGCTAGCACGTTTACTACGAGGGAATAAAAAAGAAATGAAGGAAAGCTA 16921
-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+ 16980

PCT/US05/27239 292/487

Figure 101AN

a . V L W K L A R L L R G K *
CAAACACCTTCGATCGTGCAAAATGATGCTCCCTTTATTTTCTTTACTTTCCCTTTTCGAT
16981 AGGCTGTTTCCTTTTTCGGGCTCTTTTGTCAACTGTAGTGGGTGAAAAAGCTAAGCTCG
TCCGACAAGGAAAAAGGCGGAGAAACAGTTGACATCACCCAACTTTTTCGATTCGAGC 17040
17041 AGAAAGGACAAAATTTTGTCTCTTTCTTTTGTGATATTCAGAGCGATAAAAATCCGTTTTTT
TCTTTCCTGTTTAAACAGGAAAGAAAAACATAAGTCTCGCTATTTTATTAGGCAAAAAA 17100
GAAGTTTTCAAA
17101 -----+--- 17112
CTTCAAAAGTTT

M1, strain 2580

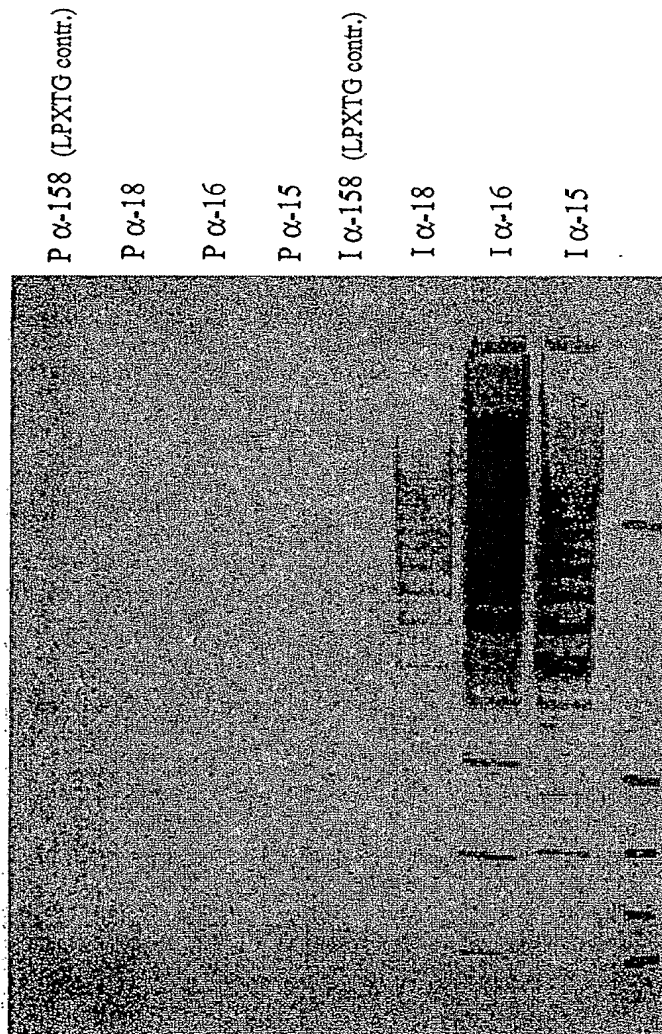


Figure 102

LEGEND:I α -#: immune serum anti-#P α -#: pre-immune serum anti-#

PCT/US2005/027239 294/487

M1, strain 2913

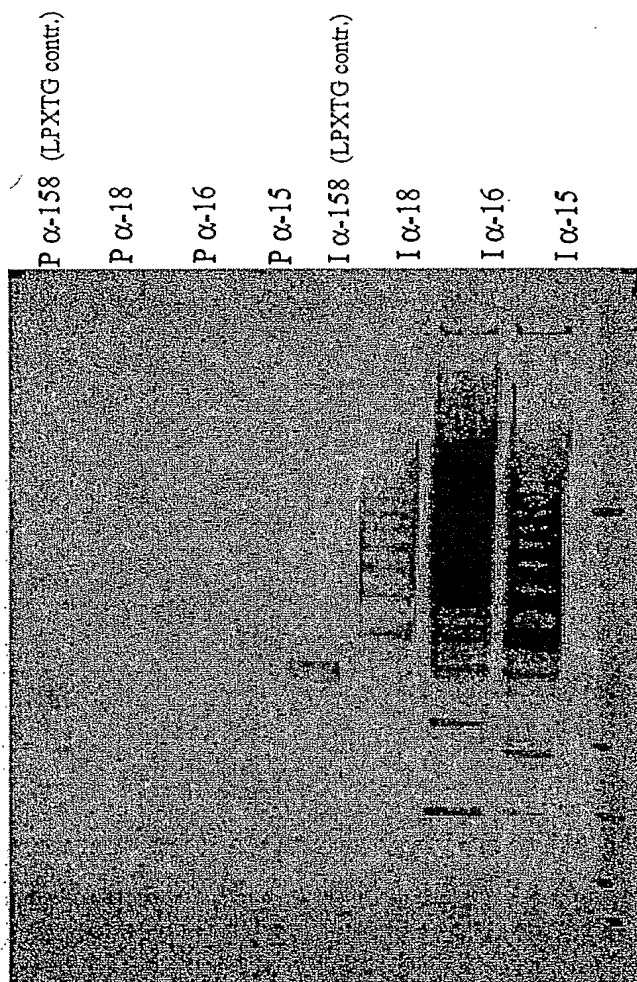


Figure 103

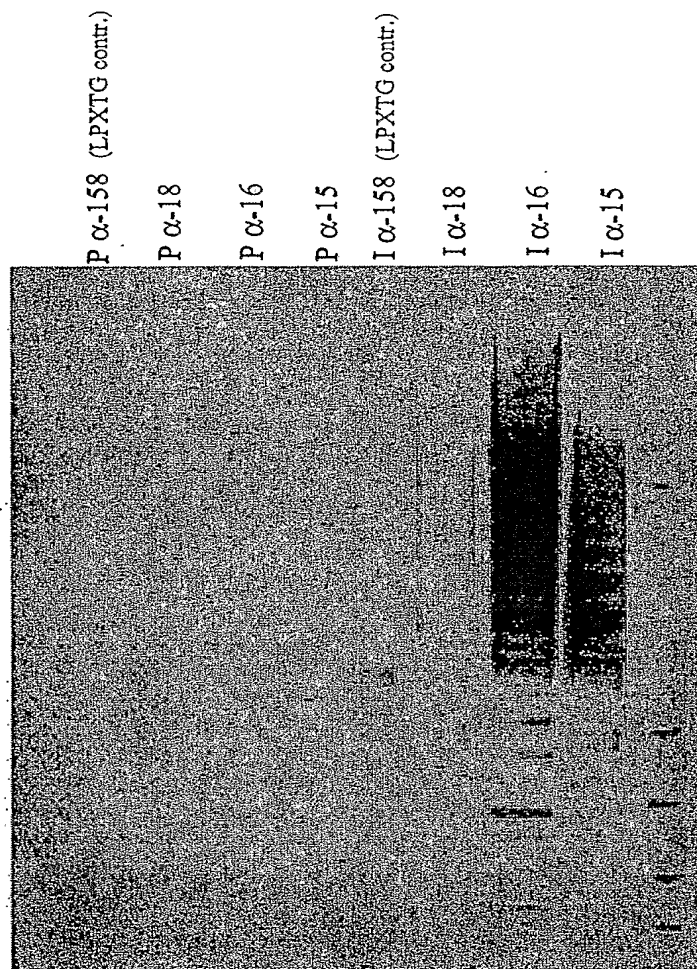
LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

295/487

M1, strain 3280

**LEGEND:**

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

Figure 104

296/487

M1 strain 3348

P α -158 (LPXTG contr.)
P α -18
P α -16
P α -15
I α -158 (LPXTG contr.)
I α -18
I α -16
I α -15

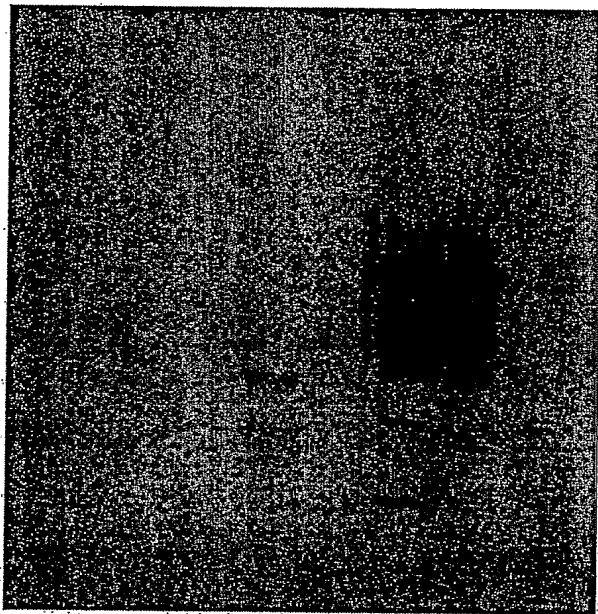


Figure 105

M1 strain 2719

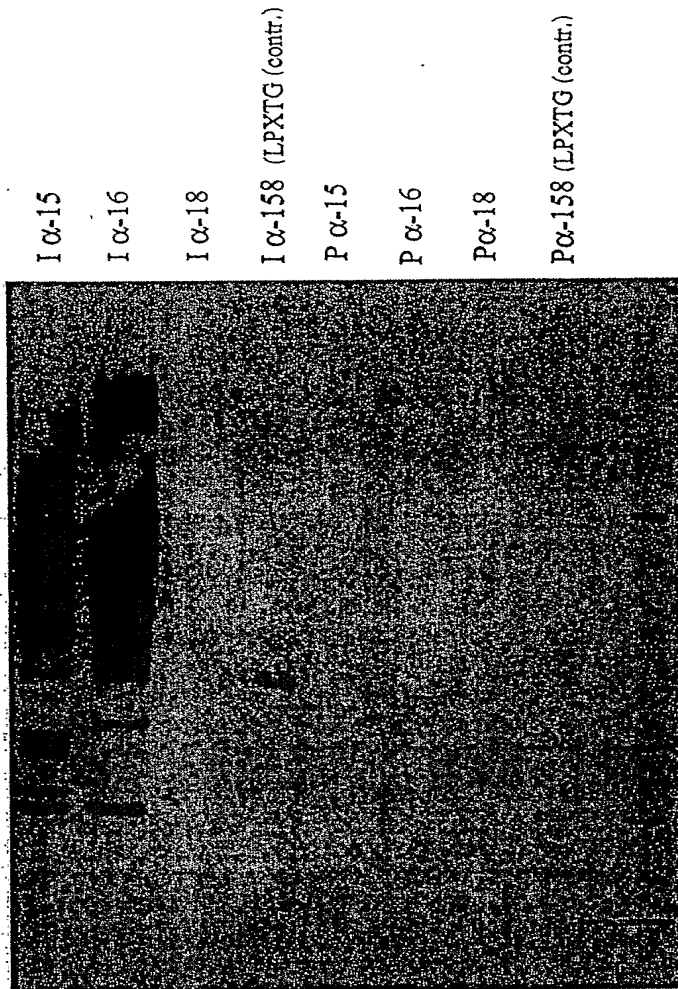


Figure 106

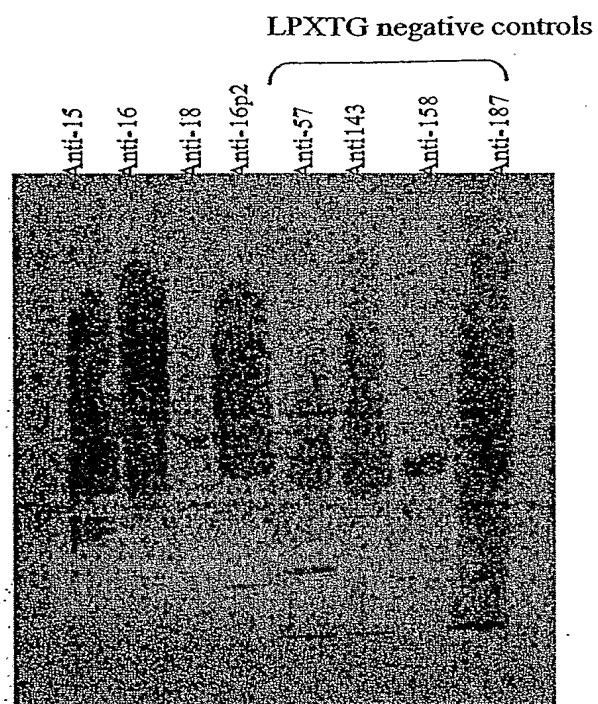
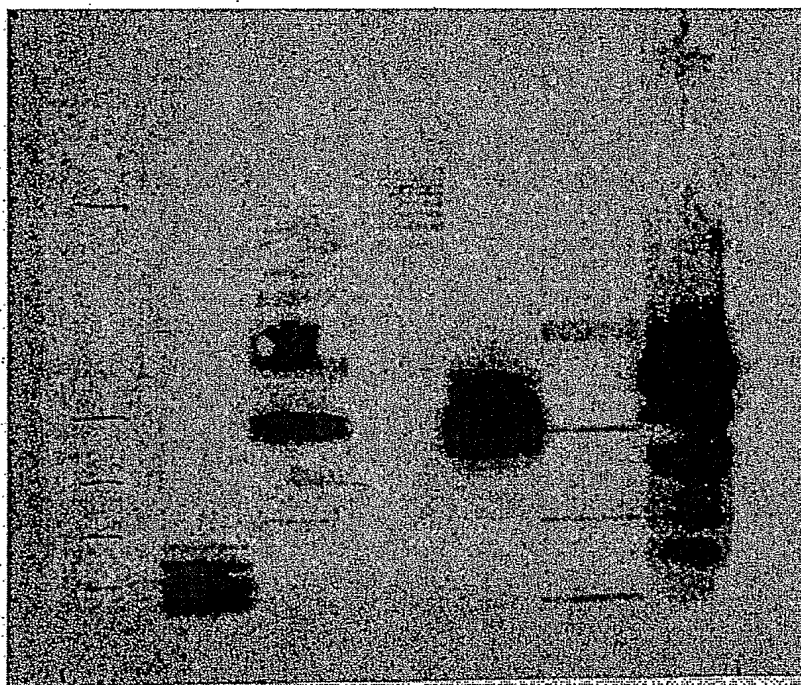
Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)

Figure 107

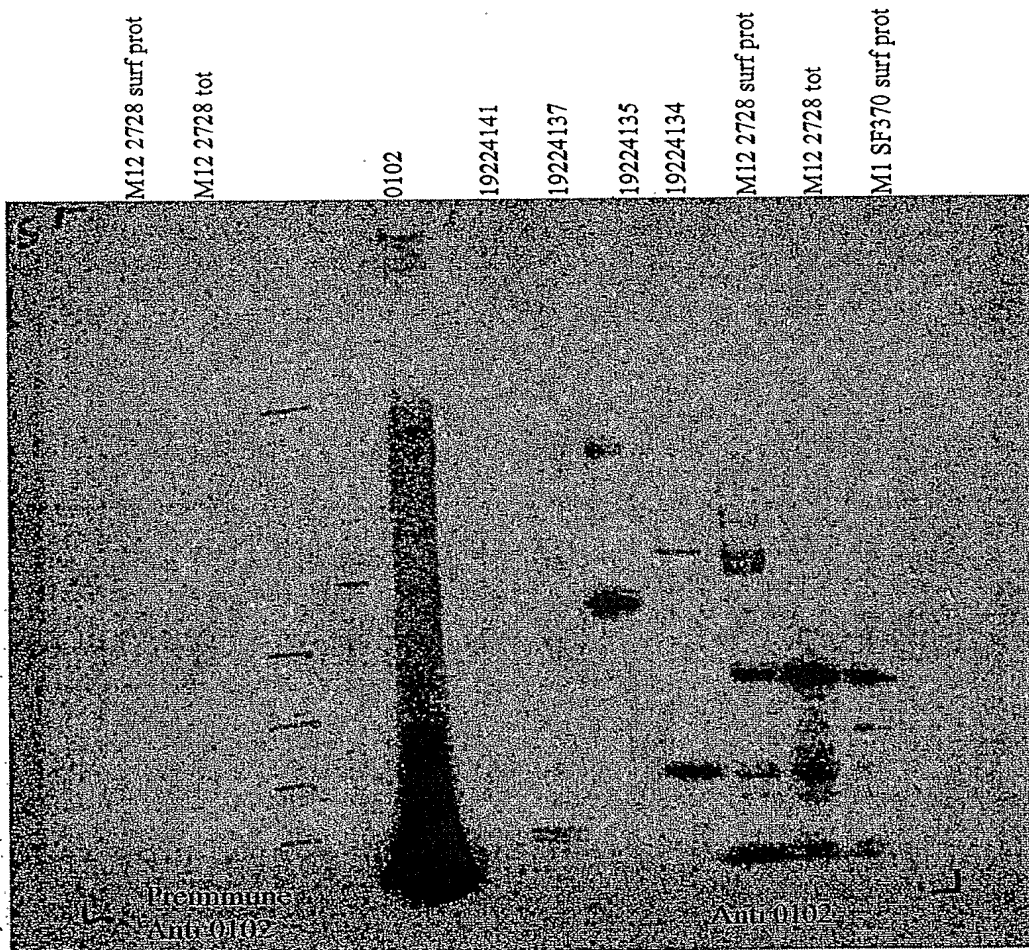
Western blot on fraction enriched in surface proteins of M12 (2728)

Anti-19224134
Anti-19224135
Anti-19224137
Anti-19224141
Anti-0102
Anti-158 (control)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

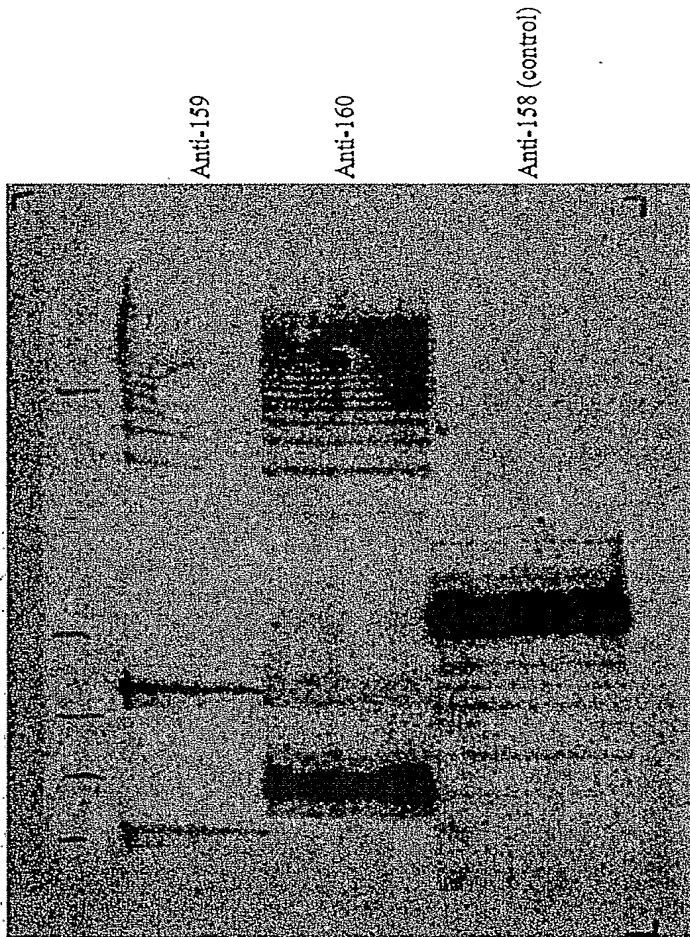
Figure 108



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

Western blot on fraction enriched in surface proteins of M6 (2724)

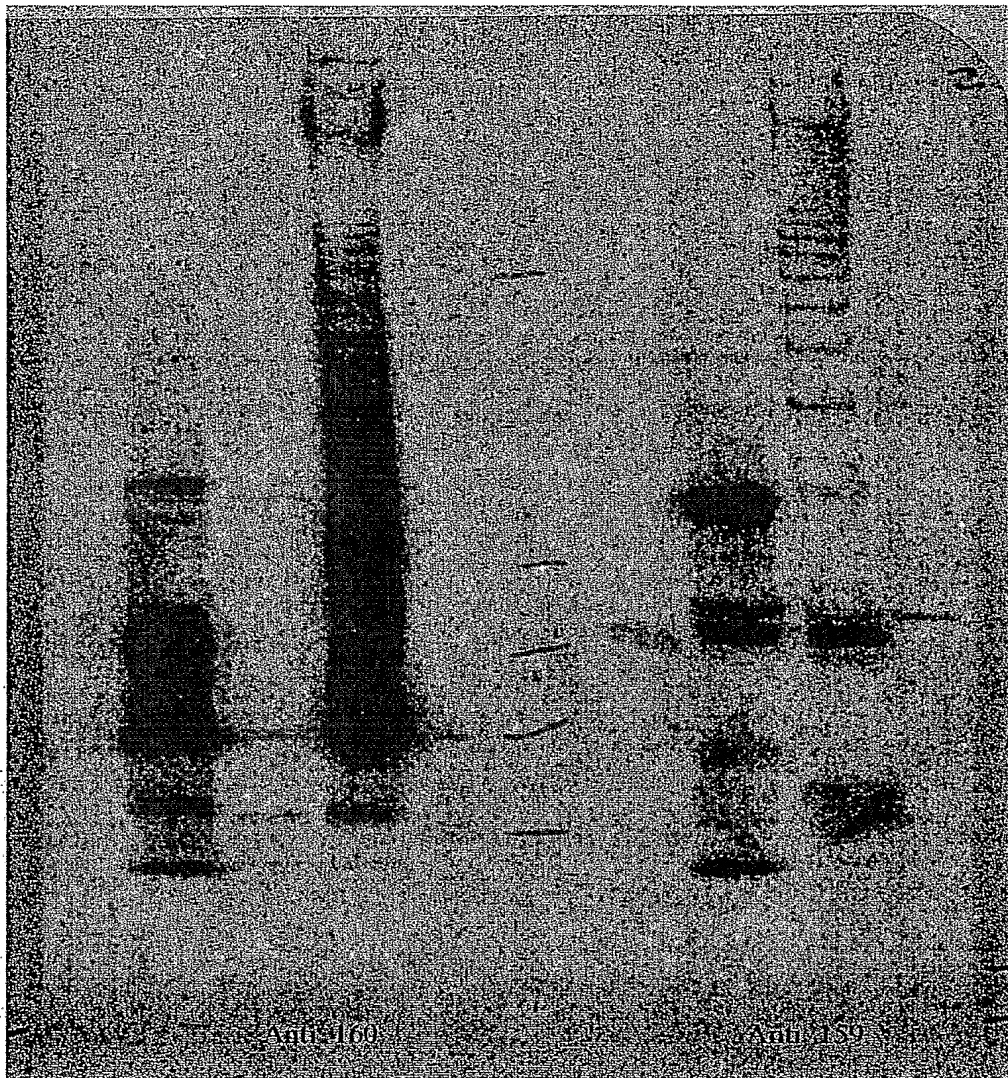


M6 strain isolate 2724

Figure 110

PCT/US05/27239

0160
0159
M6 3650 surf prot
M1 SF370 surf prot
Marker
0160
0159
M6 3650 surf prot
M1 SF370 surf prot



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

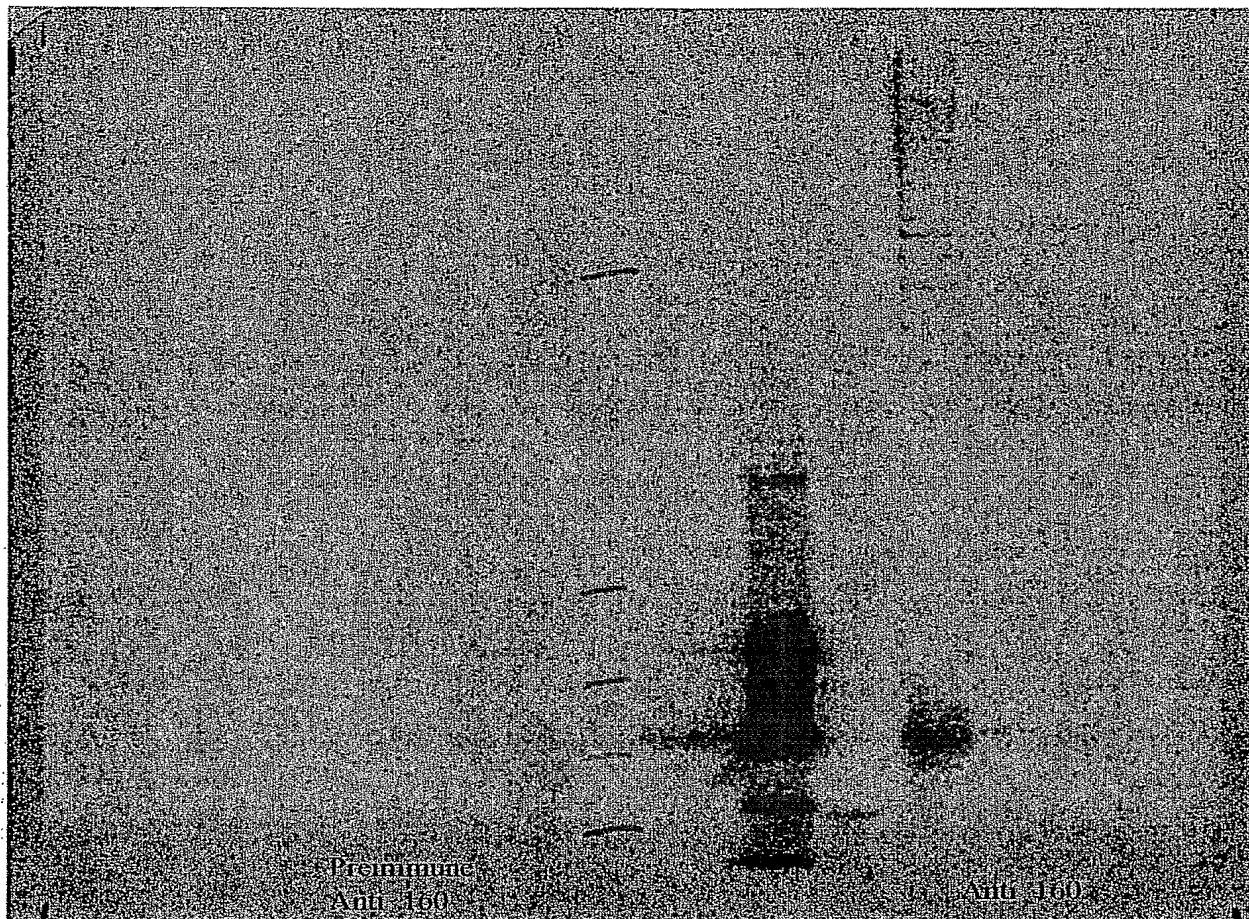


Figure 112

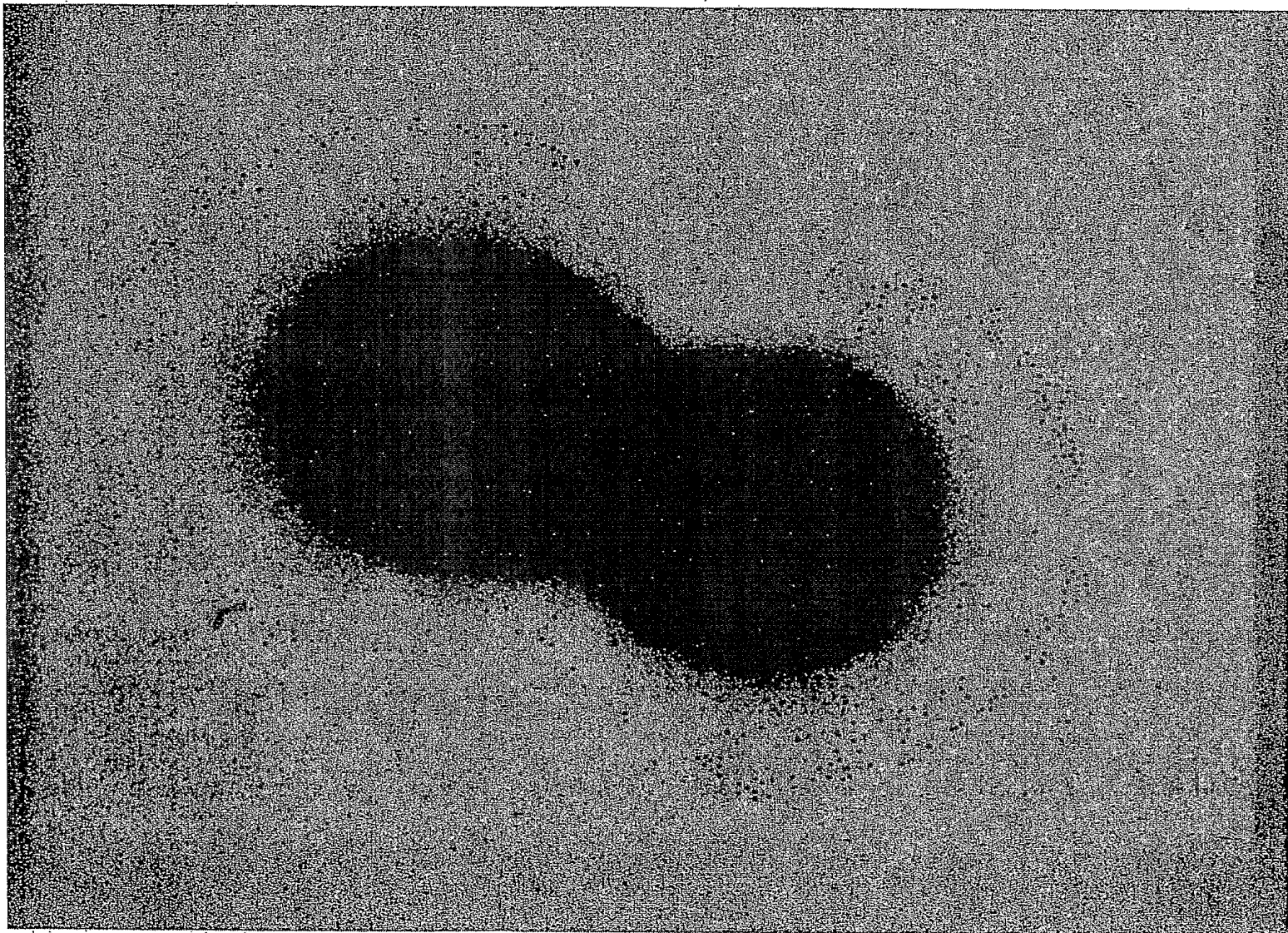


FIGURE 113

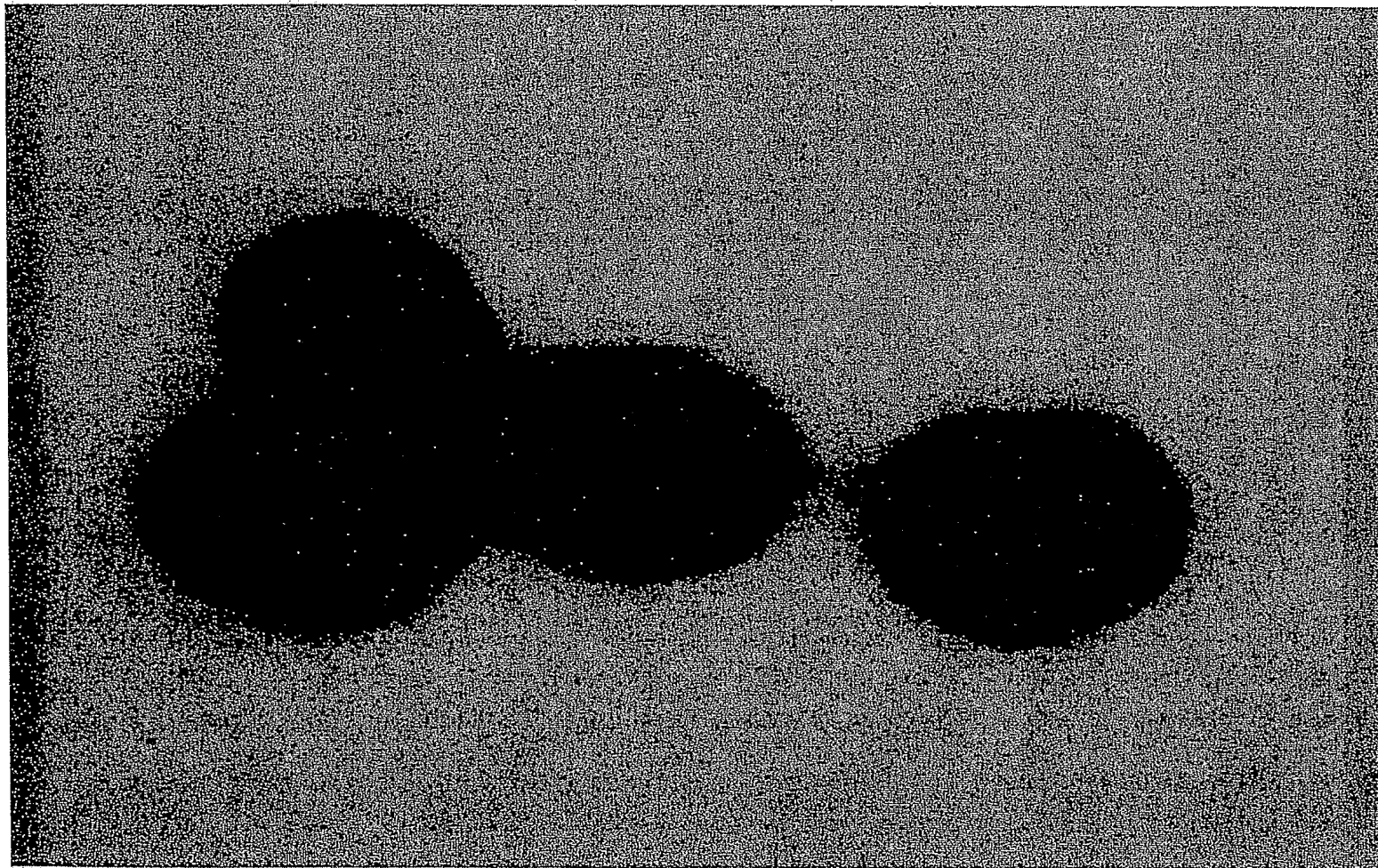


FIGURE 114

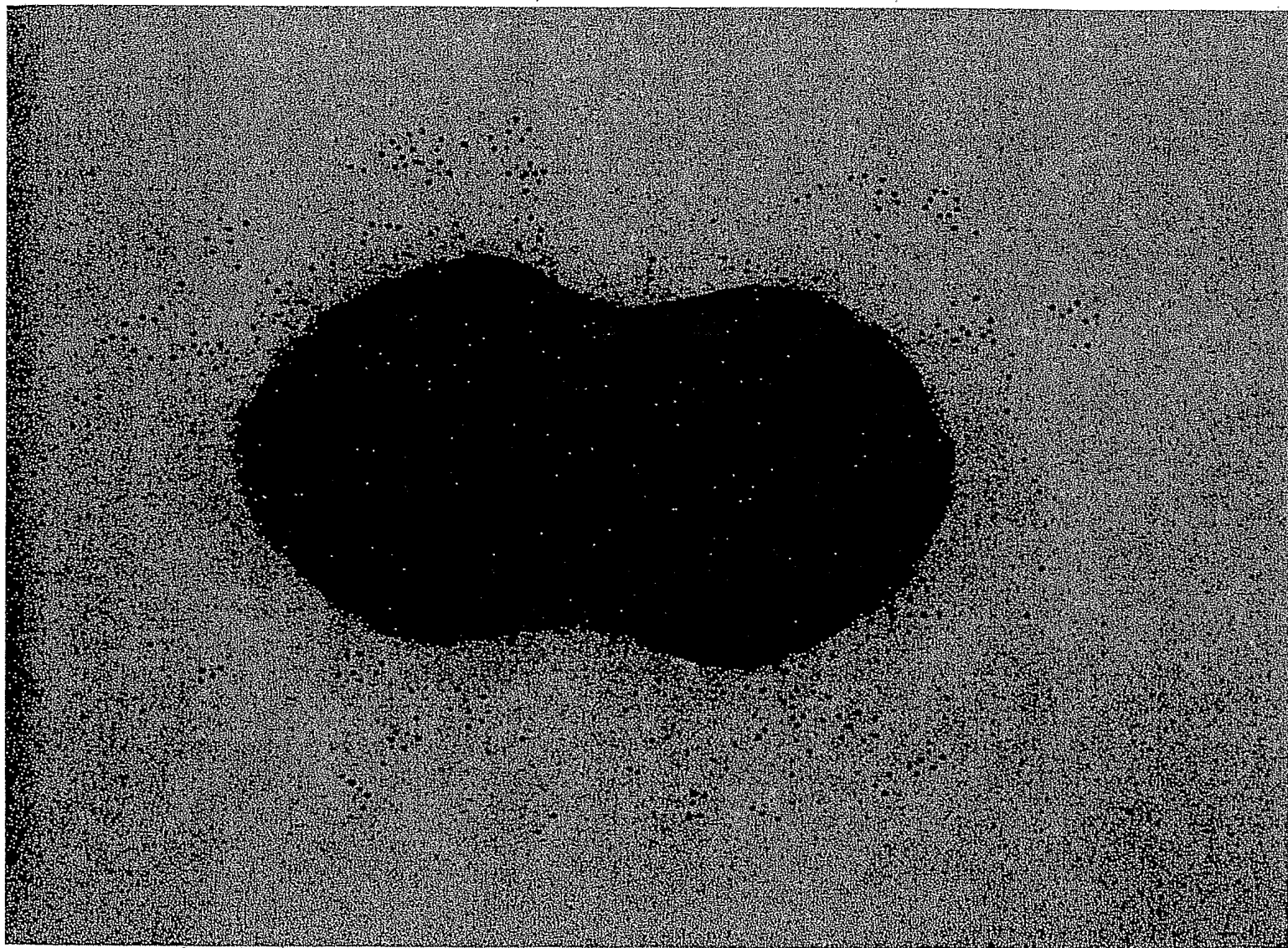


FIGURE 115

PCT/US05/27239 307/487

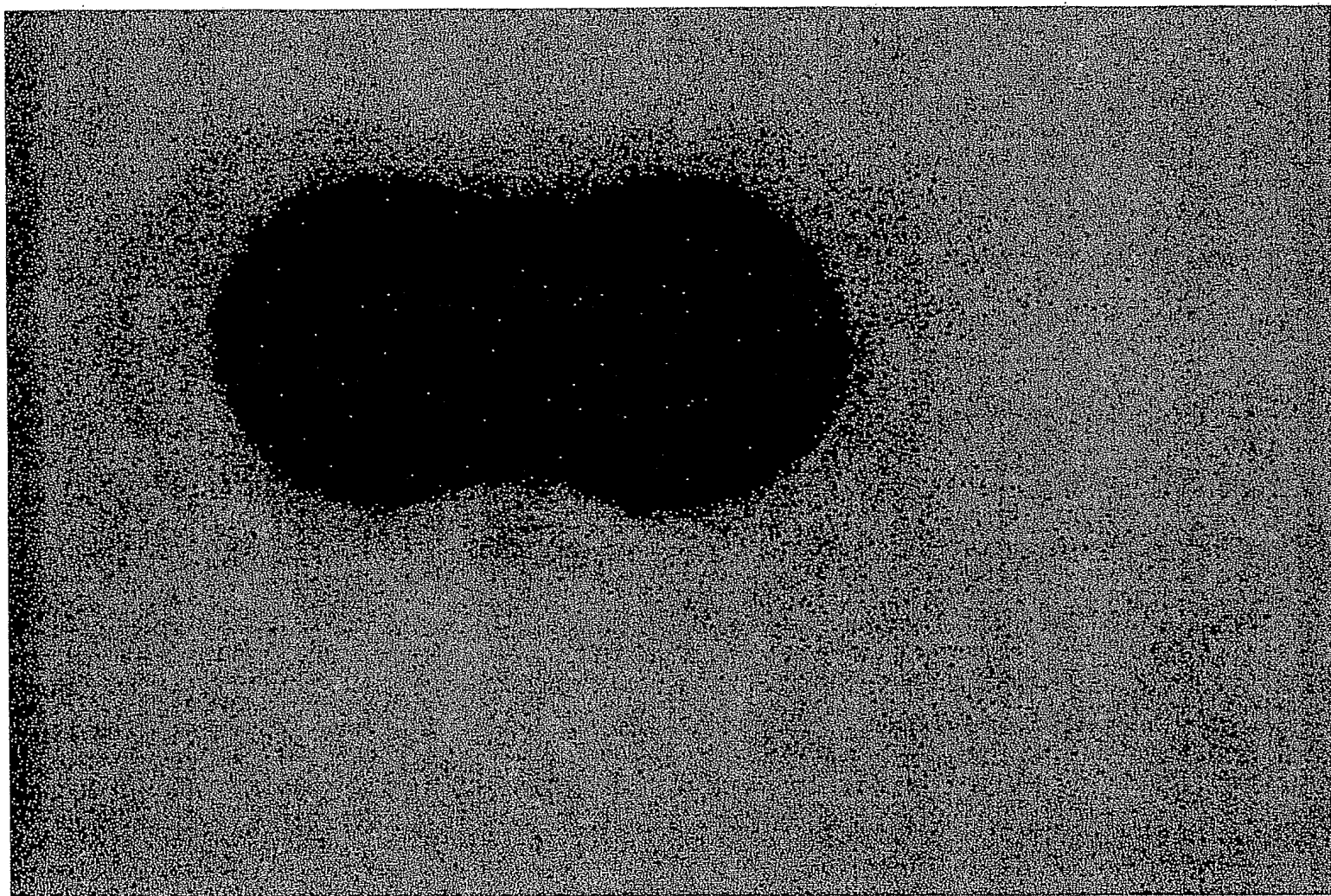


FIGURE 116

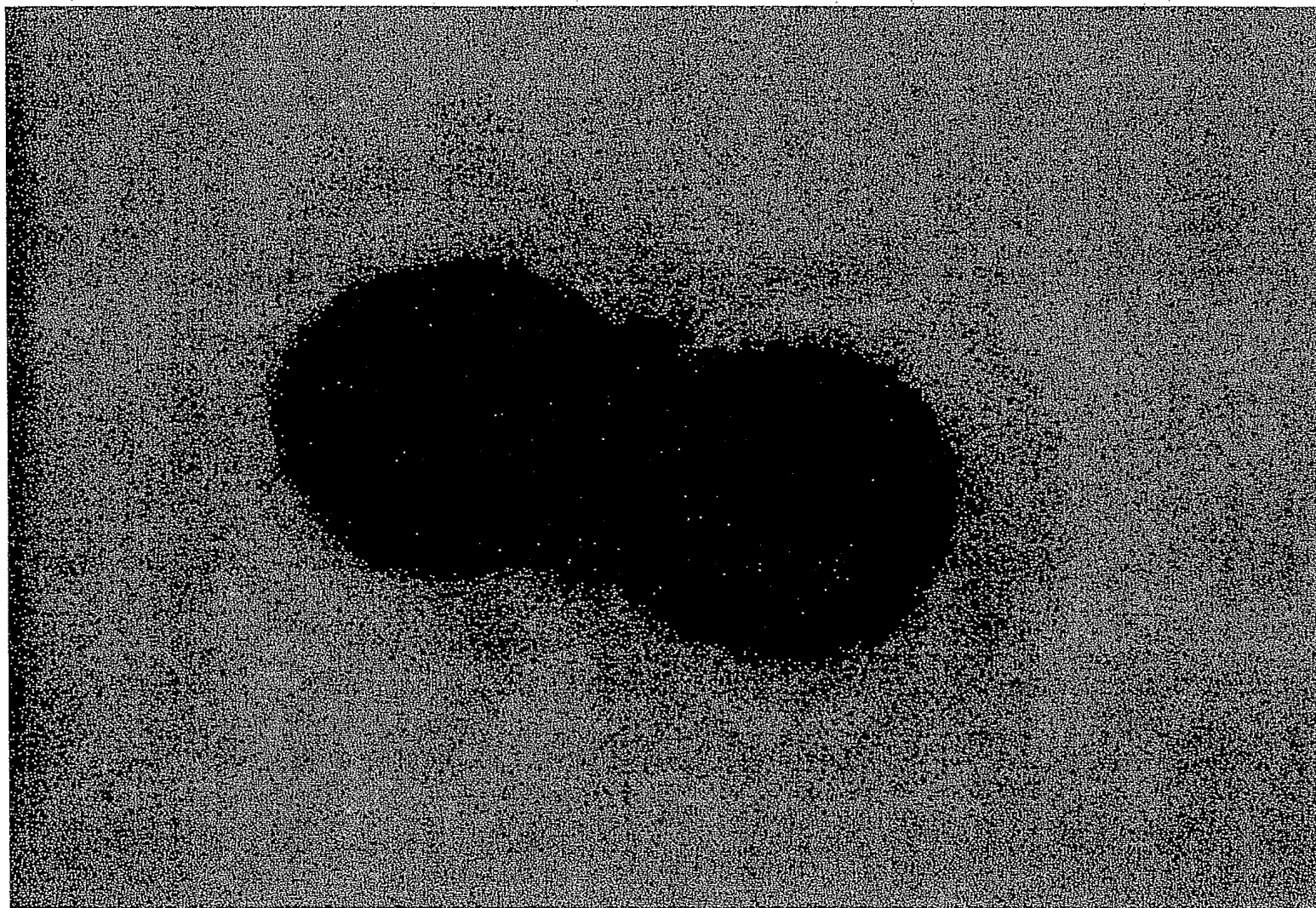


FIGURE 117

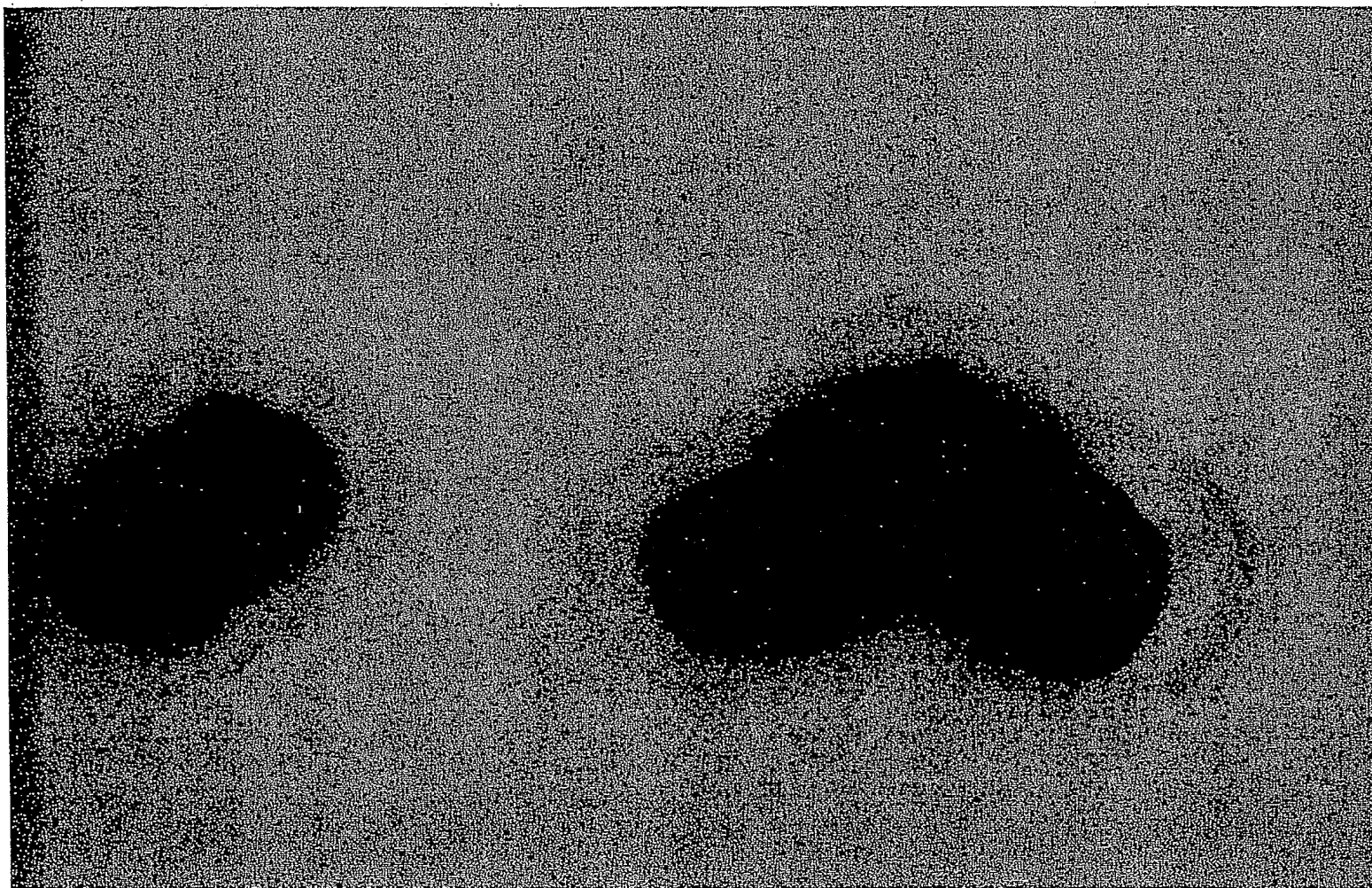


FIGURE 118

PCT/US05/27239

310/487

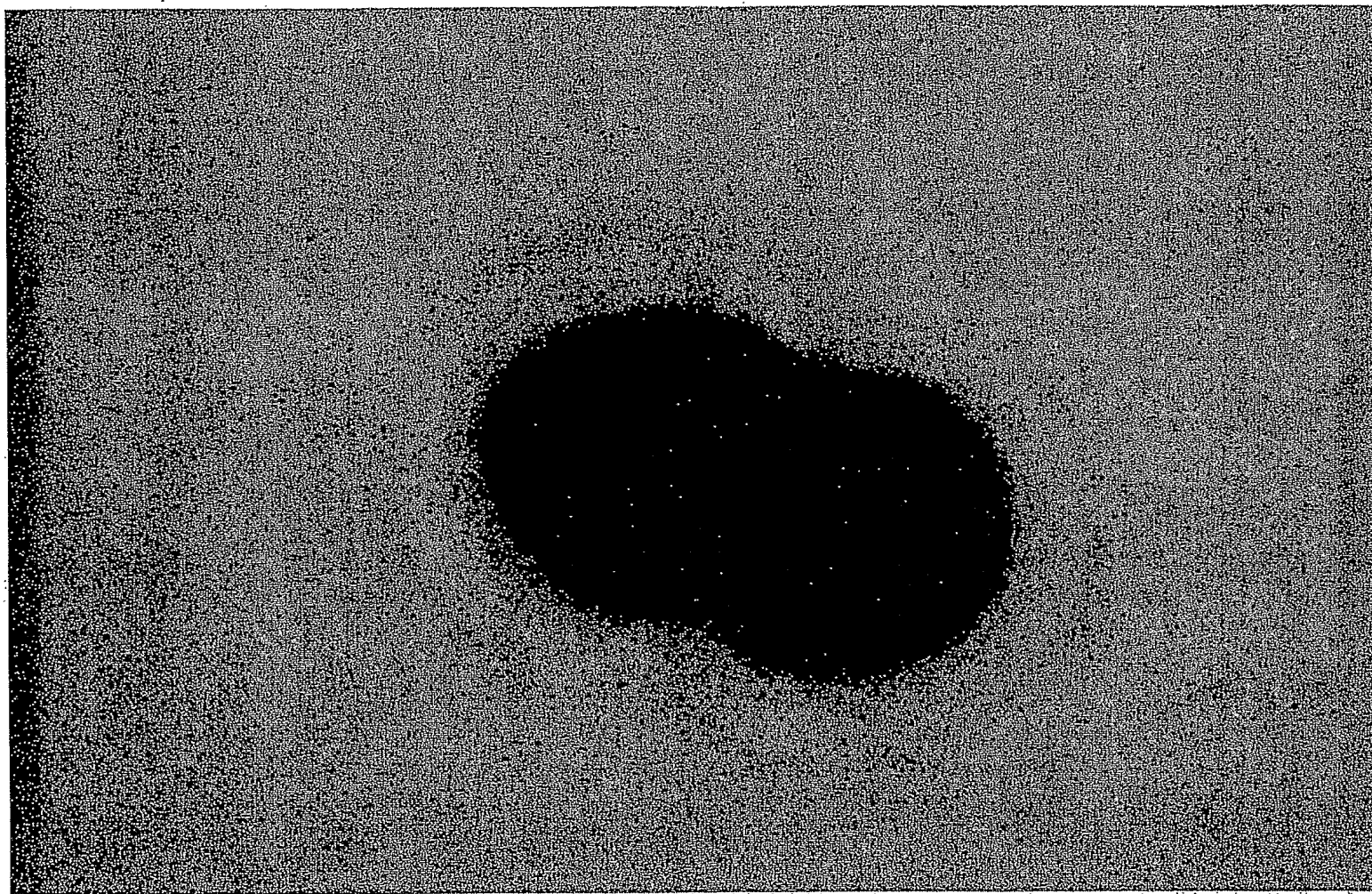


FIGURE 119

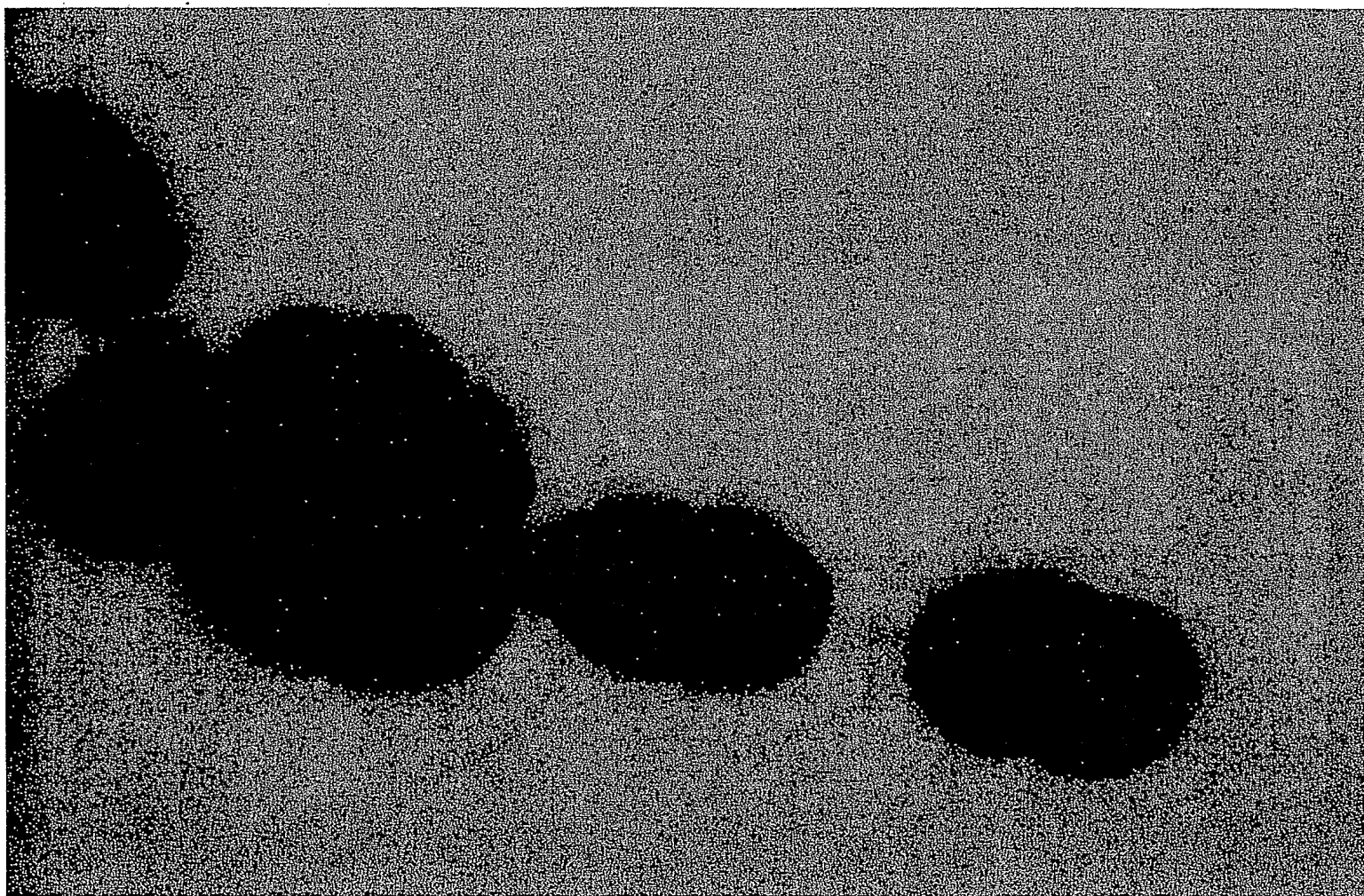


FIGURE 120

312/487

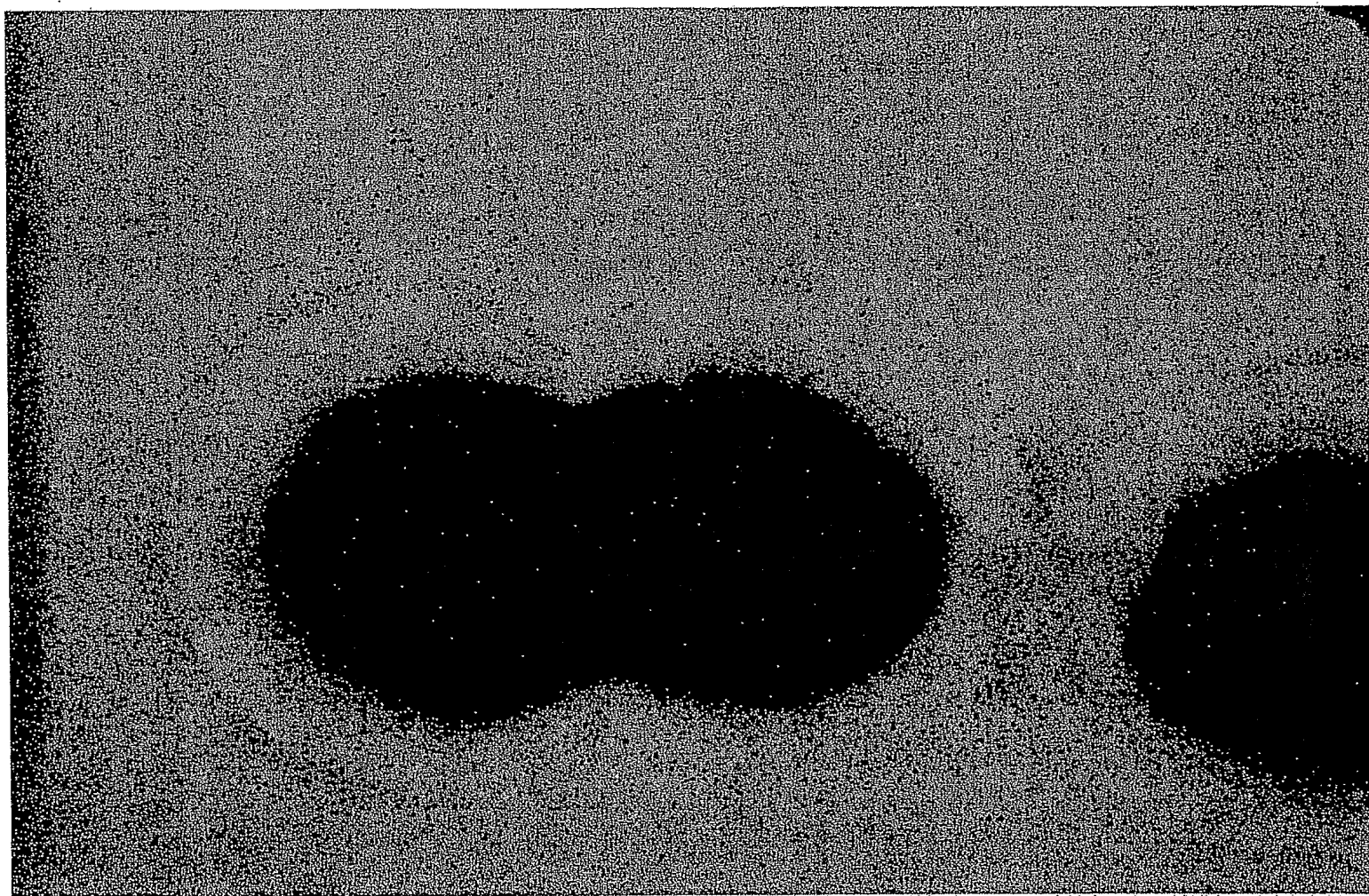


FIGURE 121

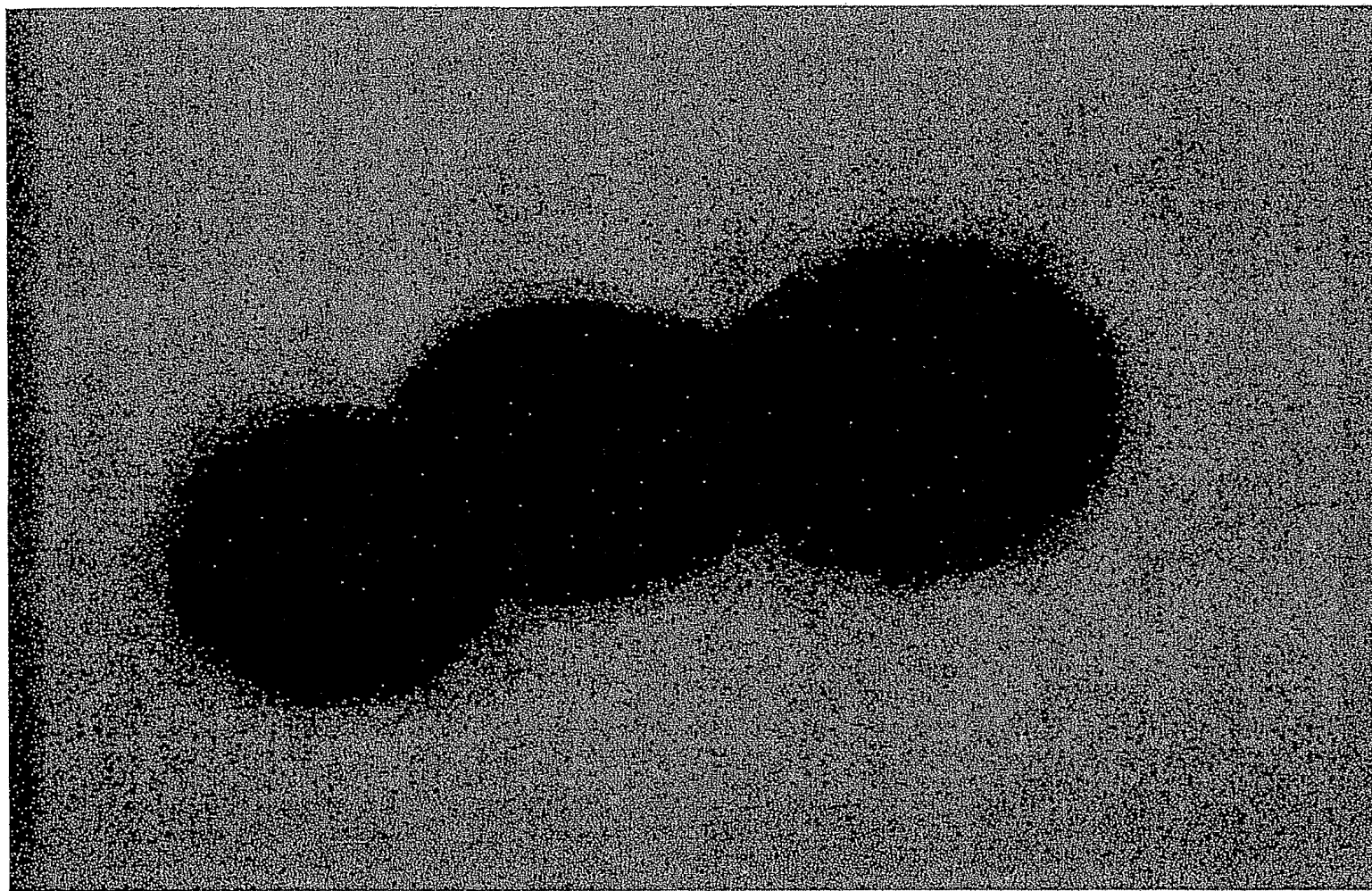


FIGURE 122

314/487

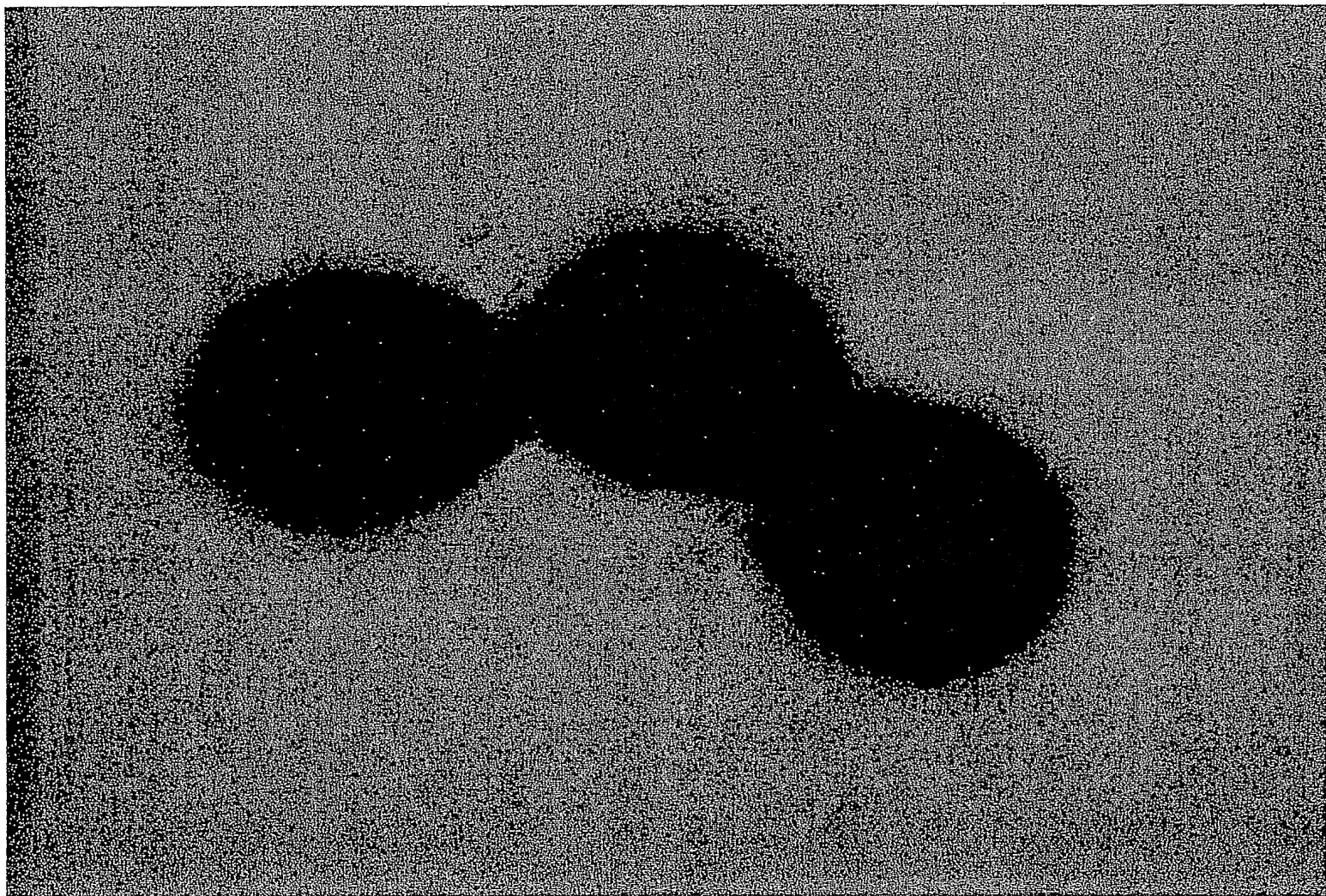


FIGURE 123

315/487

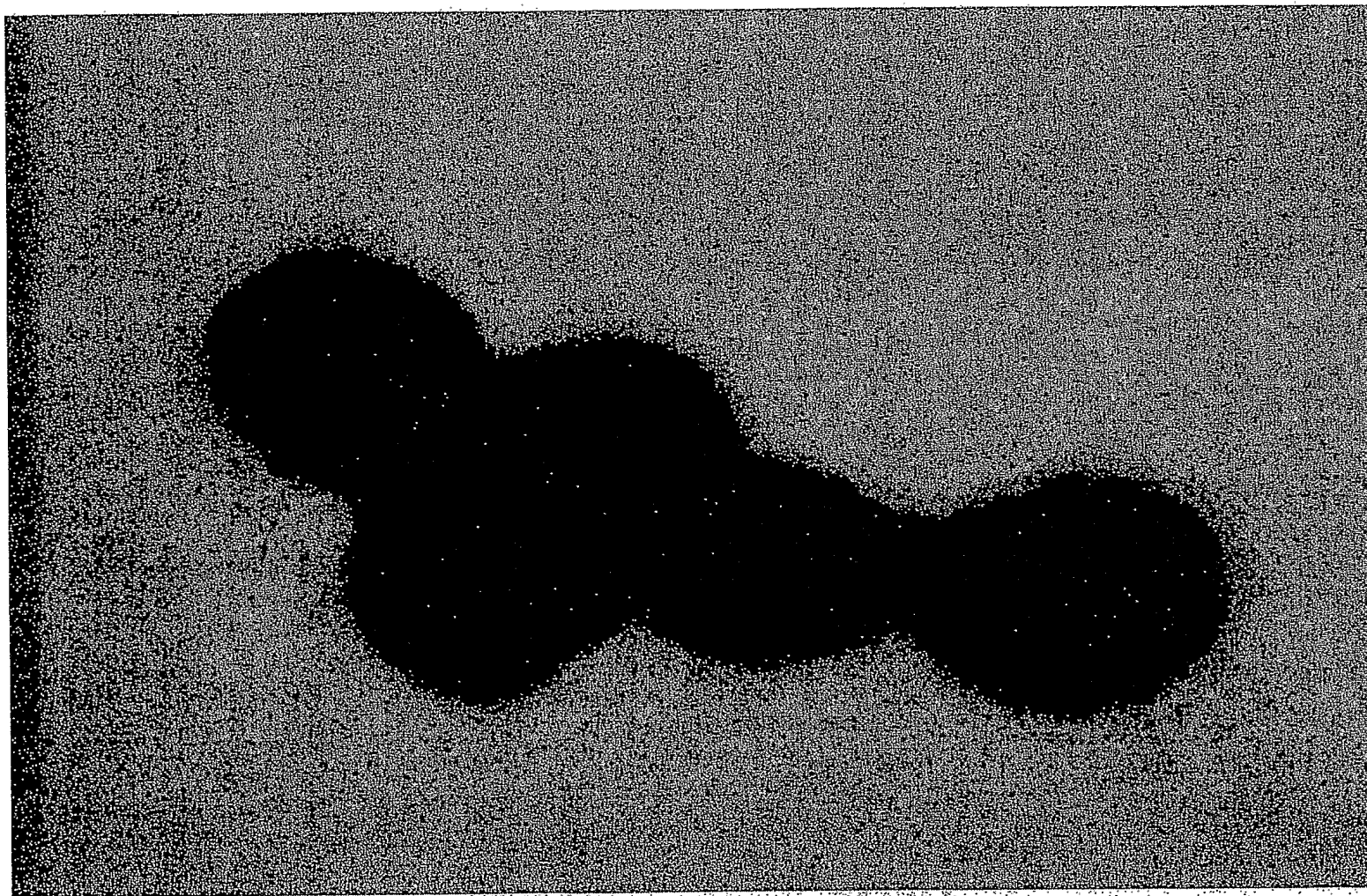


FIGURE 124

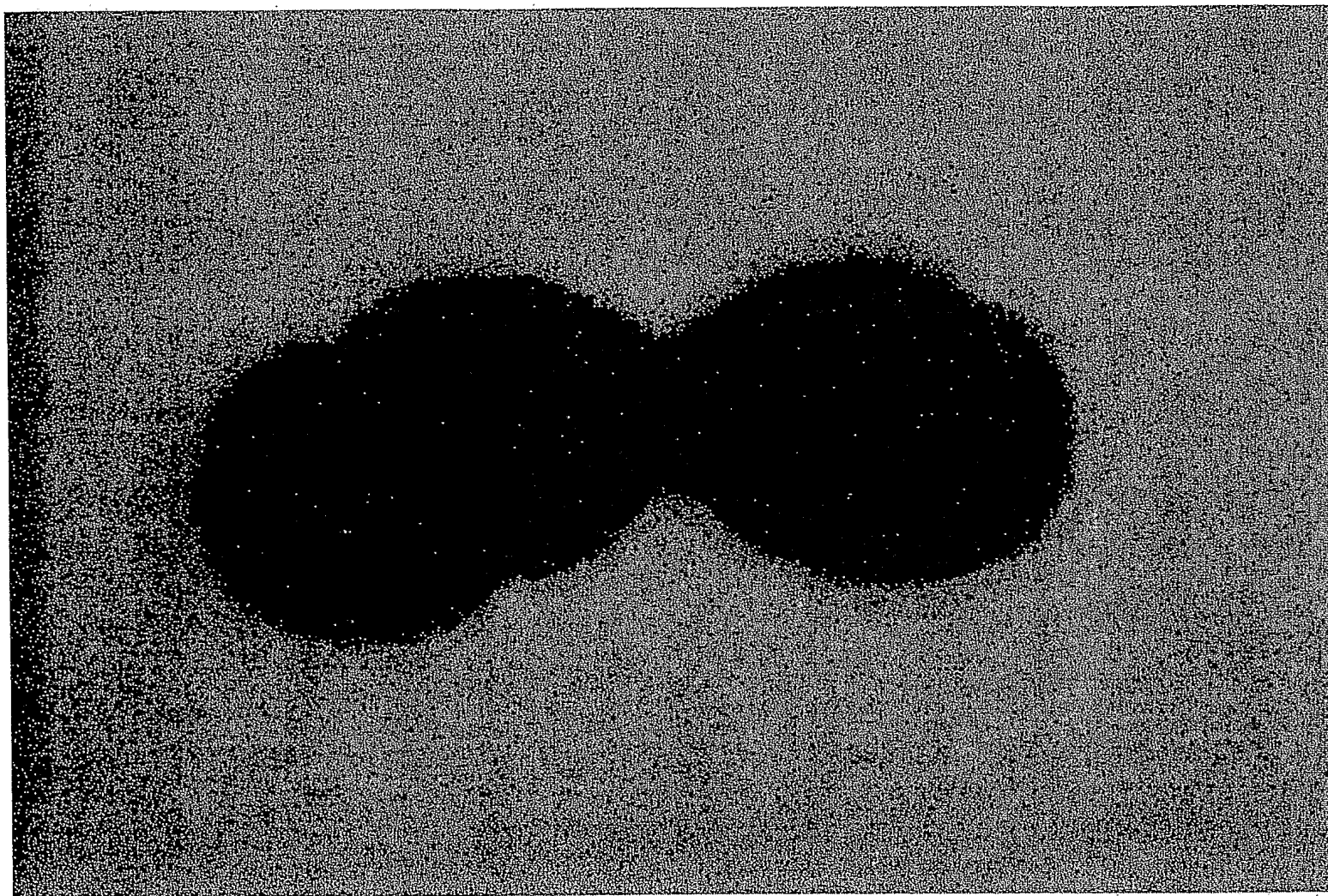


FIGURE 125

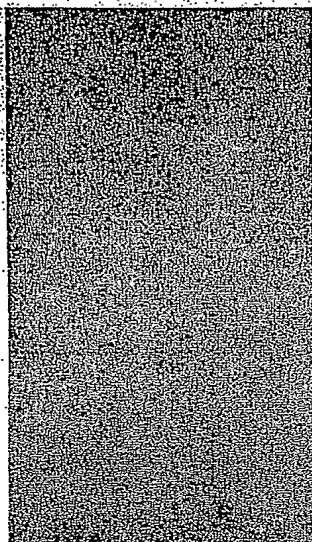
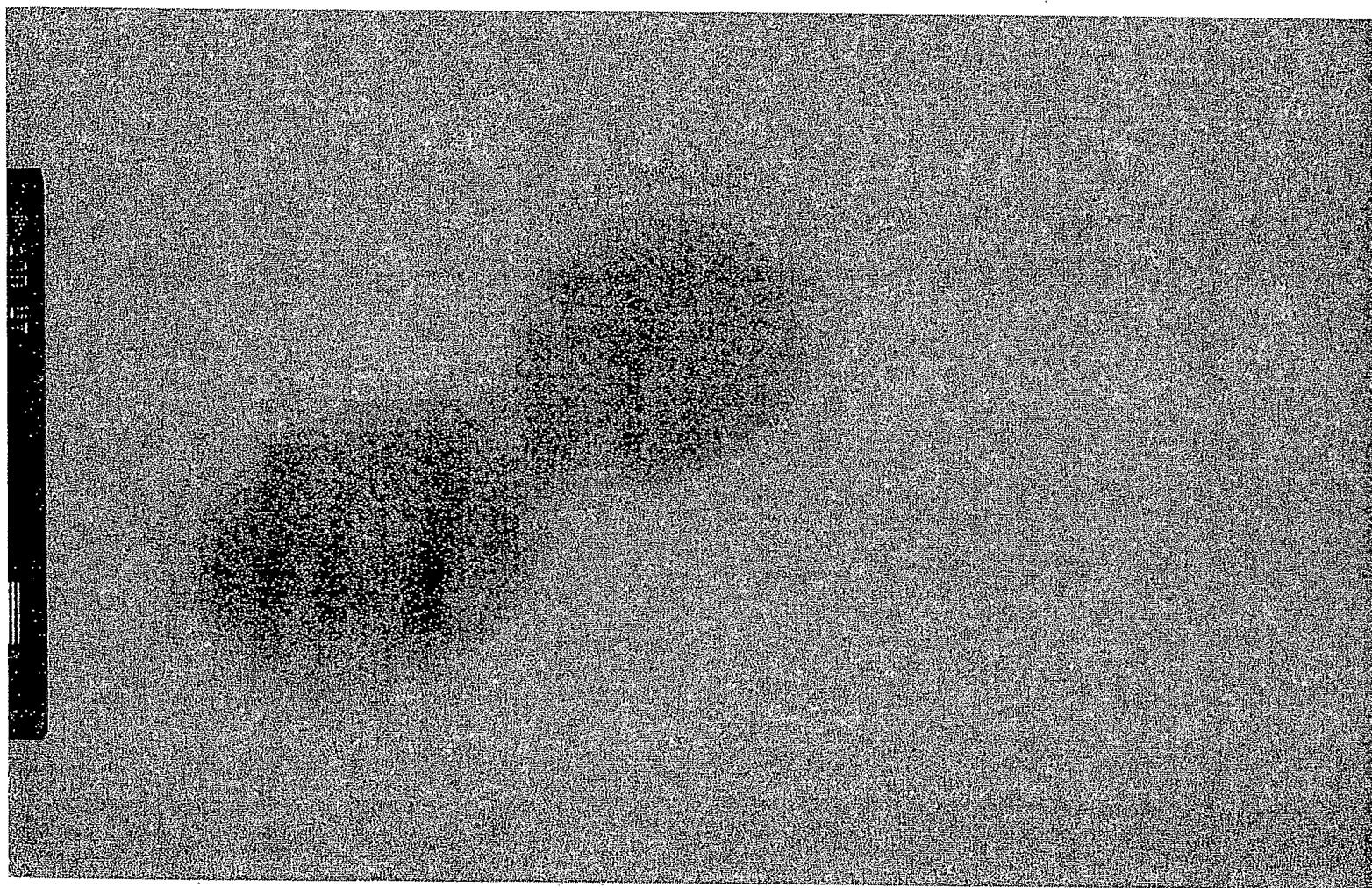


FIGURE 126

Figure 127



PCT/US05/27239 319/487

Figure 128

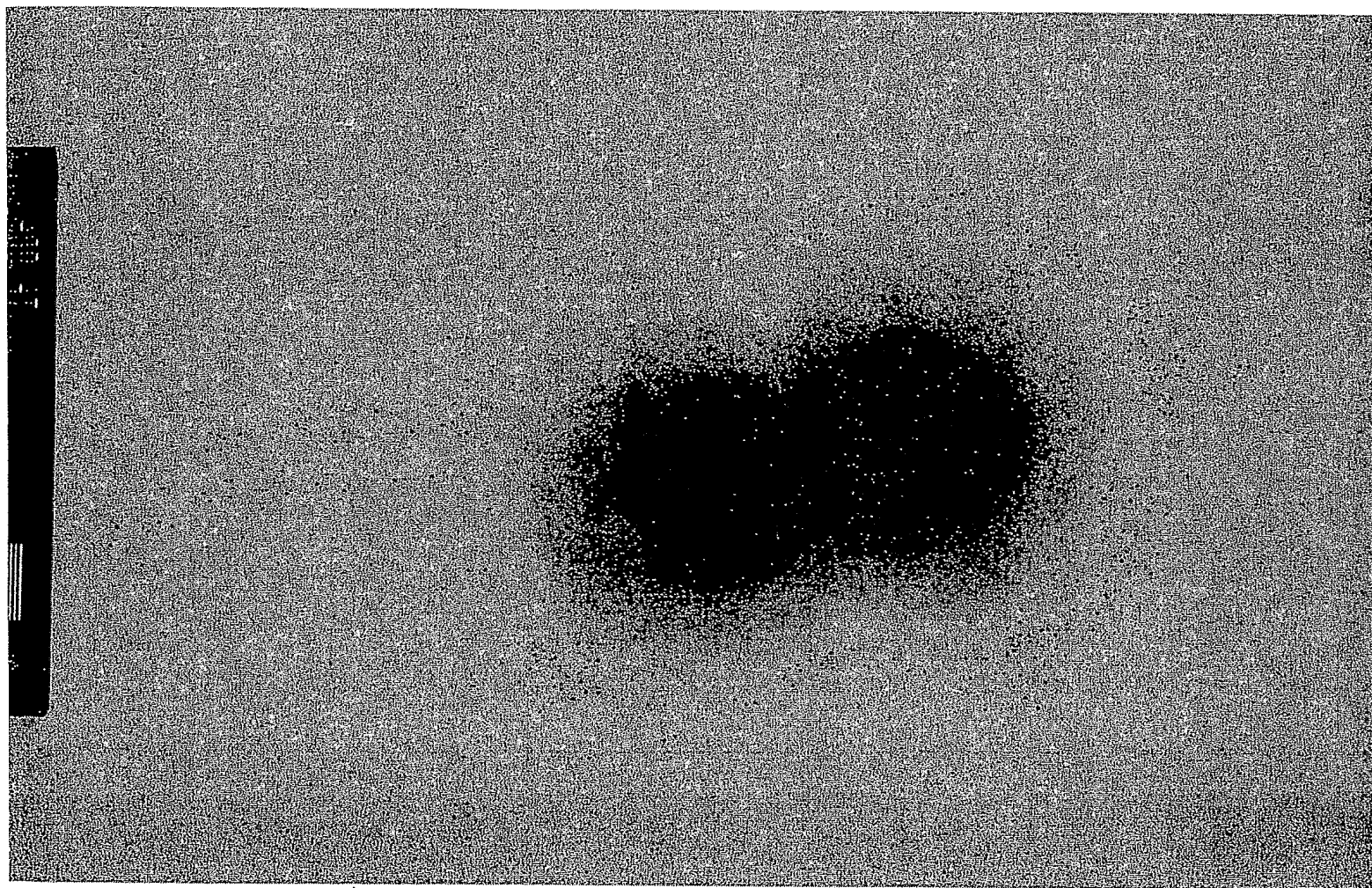


Figure 129

